Surveillance of mRNP composition during translation termination regulates gene expression via nonsense-mediated mRNA decay

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Volker Böhm

aus Bergisch Gladbach

Köln, 2015

Berichterstatter/in: PD Dr. Niels H. Gehring

Prof. Dr. Karin Schnetz

Tag der mündlichen Prüfung: 16.06.2015

Content

1. Introduction

According to the central dogma of molecular biology, access to the genetic information contained within the DNA requires the synthesis of messenger RNA (mRNA), which is decoded in order to generate proteins (Crick, 1970; Crick, 1958; Nirenberg and Matthaei, 1961). In eukaryotic cells, this process of gene expression consists of several consecutive, but integrated steps (Moore and Proudfoot, 2009). The first step is the RNA polymerase II (Pol II) mediated transcription of the DNA into pre-mRNA, which is generally accompanied by three co-transcriptional processing actions to modify the transcript (Bentley, 2014; Lee and Tarn, 2013). These involve the addition of a 7-methylguanosine cap to the 5′ end, the splicing of intronic sequences, and the cleavage at the 3′ end followed by addition of the poly(A) tail. Upon completion of these steps, the mature mRNA is exported through the nuclear pores into the cytoplasm where ribosomes translate the transcript into a polypeptide chain. Eventually the mRNA is degraded, which represents the final step in the lifecycle of an mRNA (Figure 1A) (Moore, 2005).

1.1 Surveillance of gene expression

As each of the individual processes during gene expression are carried out by specific complex machineries with inherent error rates, mistakes can occur, which need to be detected in order to prevent the generation of faulty RNA or proteins (Doma and Parker, 2007; Schmid and Jensen, 2010; Shoemaker and Green, 2012). These mistakes are, for example, the misincorporation of nucleotides or amino acids by Pol II or the ribosome during transcription or translation, respectively. Moreover, different stimuli, chemical agents or environmental influences can increase the error frequency of these processes, thereby potentially producing more aberrant gene expression products (Drummond and Wilke, 2009; Jack et al., 2011; Remenyi et al., 2004; Wurtmann and Wolin, 2009; Zaher and Green, 2009). Eukaryotic cells employ several quality control mechanisms at basically every nuclear and cytoplasmic gene expression step in order to detect abnormalities (Figure 1B) (Ghosh and Jacobson, 2010; Muhlemann and Jensen, 2012). Although failsafe mechanisms are also implemented in the gene expression machineries themselves, quality control commonly involves the handover of only correctly processed products from one step to the other, whereas faulty intermediates are retained or degraded (Doma and Parker, 2007; Hagiwara and Nojima, 2007; Maniatis and Reed, 2002). On the molecular level, this is achieved by the interplay of specific proteins, which bind

the mRNA to form ribonucleoproteins (RNP) (Muller-McNicoll and Neugebauer, 2013; Rodriguez-Navarro and Hurt, 2011). The composition of the mRNP changes dramatically during the progression of gene expression and determines the fate of the transcript (Figure 1A) (Mitchell and Parker, 2014; Singh et al., 2015).

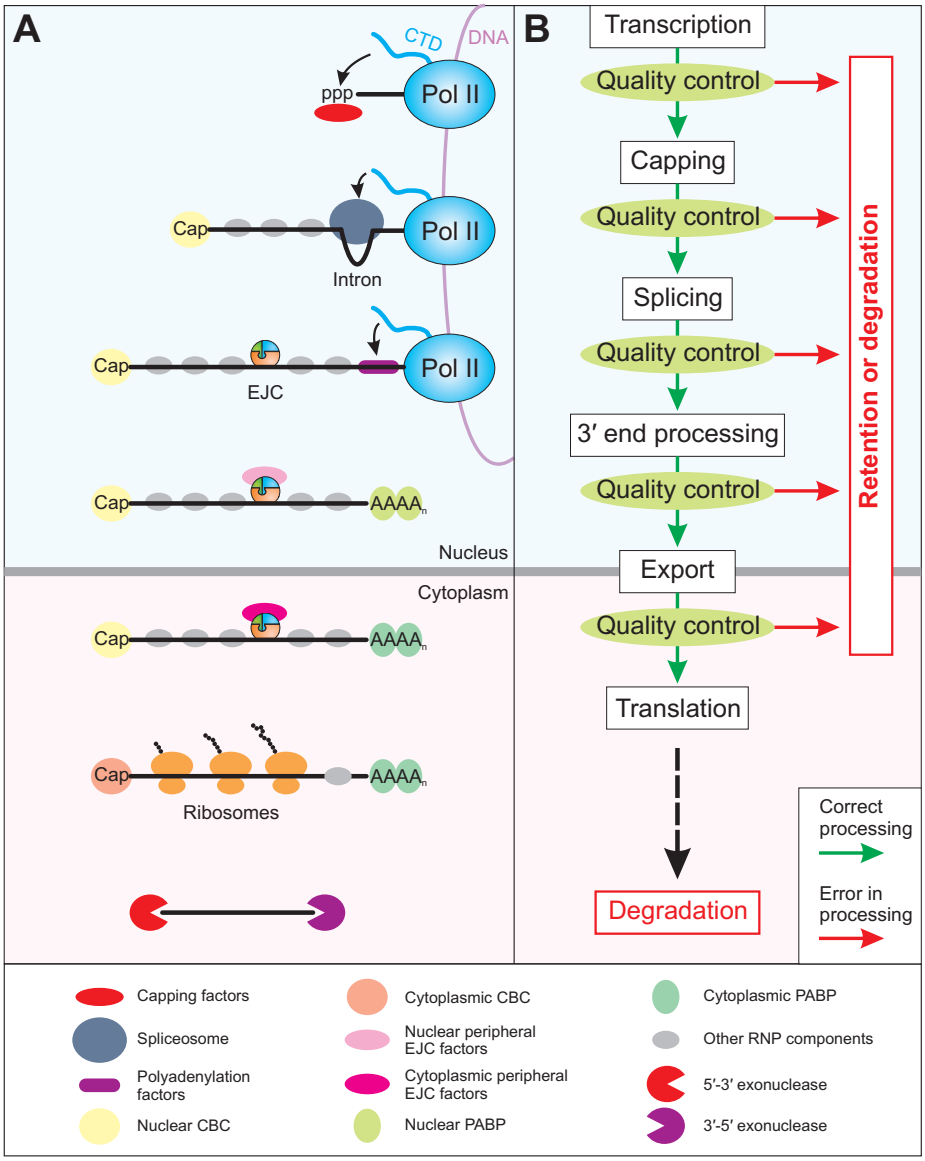


Figure 1: Overview of mRNP composition and quality control during gene expression. (A) Central steps of gene expression are depicted schematically. mRNA processing factors are recruited to the mRNA co-transcriptionally by the C-terminal domain (CTD) of Pol II. The Legend of mRNP components is shown at the bottom (A). (B) The fidelity of each step of the gene expression cascade (white boxes) is monitored by quality control mechanisms, which initiate the degradation or retention of erroneous product. Only correctly processed or quality control evading gene expression intermediates are handed over to the next step (green arrows). Finally, the mRNA is translated in the cytoplasm and eventually degraded. Abbreviations: Pol II = RNA polymerase II; ppp = triphosphate; EJC = exon-junction complex; CBC = cap-binding complex; PABP = Poly(A) binding protein.

In metazoan cells, one key regulator of gene expression is the exon-junction complex (EJC), which is a well-studied example of a multi-protein complex that shapes the mRNP and influences many subsequent gene expression steps. EJCs are deposited in the nucleus on spliced mRNAs closely upstream of the exon-exon junction (Figure 1A) (Le Hir et al., 2000). They

remain associated with the mRNP and serve as a molecular memory of splicing until they are displaced by the translating ribosome in the cytoplasm (Gehring et al., 2009b). The core of the EJC is comprised of the DEAD-box helicase eIF4A3, Barentsz (BTZ; also MLN51) and the heterodimer Y14/MAGOH. Of the core EJC proteins, ATP-loaded eIF4A3 directly binds the RNA in a sequence-independent manner due to interaction with the phosphate-sugar backbone (Andersen et al., 2006; Bono et al., 2006). Interaction with the RNA stimulates the hydrolysis of ATP, which leads to dissociation of eIF4A3. To lock the EJC stably on the RNA, the Y14/MAGOH dimer binds to RNA-interacting eIF4A3 and keeps it in a state that prevents the completion of ATP hydrolysis (Ballut et al., 2005). Specific disassembly of EJCs is achieved by ribosome-associated PYM (partner of Y14 and MAGOH), which lifts Y14/MAGOH from the EJC resulting in the release of eIF4A3 from the RNA (Bono and Gehring, 2011).

At certain steps during the mRNP metabolism, additional EJC proteins can join and leave the core factors (Bono and Gehring, 2011; Tange et al., 2004). By recruitment of specific protein factors, the dynamic composition of the EJC changes and allows for the activation of downstream processes in the mRNP lifecycle, such as mRNA export or translation (Chazal et al., 2013; Gudikote et al., 2005; Le Hir et al., 2001; Nott et al., 2004; Wiegand et al., 2003). Therefore, the EJC is not only an example for the tight interplay of the gene expression processes, it also represents an integral component of mRNP quality control, because only correctly spliced mRNAs benefit from the enhancing effects of the EJC.

1.2 Translation-coupled mRNP quality control

The mRNPs being translated in the cytoplasm have gone through multiple controlled processing steps and are therefore supposed to contain the proper mRNA modifications and mRNP composition needed for the synthesis of functional protein. Major errors should have been corrected at this point, for example newly transcribed mRNA, which fail to be correctly 5′-capped, will be degraded or retained in the nucleus until properly processed (Doma and Parker, 2007). However, not all potentially occurring errors can be recognized and corrected by the quality control mechanisms, especially if the mistakes are subtle. For instance, nucleotide misincorporations during transcription by Pol II, which evade the inherent proofreading mechanism, are difficult to detect by the downstream surveillance machineries (Li et al., 2011).

One of the frequent and subtle errors that occur during gene expression is the acquisition of premature translation termination codons (PTC) (Savas et al., 2006). The presence of a PTC in

the mRNA leads to the abortion of protein synthesis before the complete open reading frame has been translated and, therefore, results in the synthesis of truncated, non-functional or even harmful proteins (compare Figure 2A and B) (Frischmeyer and Dietz, 1999; Holbrook et al., 2004). Possible ways to generate PTCs on the DNA level are somatic rearrangements, nonsense mutations, as well as deletions and insertions that shift the reading frame. Furthermore, mutations in functional elements or motifs such as splice sites or splicing regulatory sites can lead to differently spliced, PTC containing transcripts (Nicholson et al., 2010). PTCs can also arise on the RNA level by transcription errors (incorporation of incorrect bases or loss of register) or alternative splicing events (e.g. intron inclusion or exon skipping). It has been calculated that about one-third of all alternative splicing events in human multi-exon genes result in PTC containing mRNA (Lewis et al., 2003).

Early studies discovered that the truncated proteins encoded by the PTC-transcripts are not efficiently produced, but that the mRNA itself is degraded (Chang and Kan, 1979). Active translation is required for this process, as the presence of translation inhibiting antibiotics or stable secondary structures in the 5′ UTR of the PTC-containing mRNA result in increased PTC-mRNA levels (Belgrader et al., 1993; Carter et al., 1995). This implicates that a translation-coupled surveillance system monitors the identity of the stop codon and decides whether the ribosome stalls at a normal or a premature termination codon. This mechanism was termed nonsense-mediated mRNA decay (NMD) and represents one of the three characterized translation-dependent mRNA quality control systems (Figure 2) (Shoemaker and Green, 2012).

The other pathways, non-stop decay (NSD) and no-go decay (NGD), detect and degrade mRNAs lacking a termination codon or containing strong ribosome stalling sites, respectively (Figure 2C and D) (Isken and Maquat, 2007; Wilson et al., 2008). These two systems are more similar to each other, compared to NMD, since they do not terminate translation upon encountering a stop codon. Moreover, NGD and NSD utilize the same factors for recognition and clearance of the erroneous transcript, whereas NMD relies on different proteins.

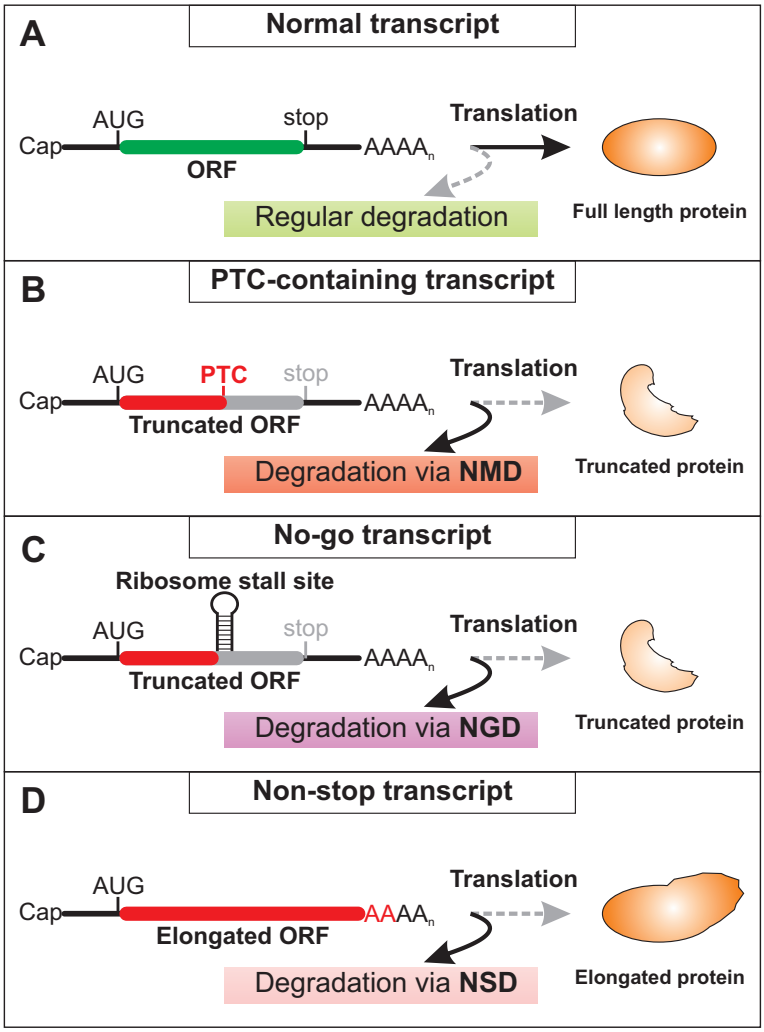


Figure 2: Comparison of translation-coupled quality control mechanisms. (A) Normal mRNAs are translated and give rise to full length, functional proteins. These mRNAs are not targeted for accelerated degradation in a translation-dependent manner. (B) The presence of a premature translation termination codon (PTC) disrupts the open reading frame (ORF) and results in shortened proteins upon translation. PTC-containing mRNA are removed by nonsense-mediated mRNA decay (NMD) during translation. (C) Strong secondary structures or other components of the mRNP can stall ribosomes upstream of the termination codon. This results, similarly to (B), in the potential generation of truncated protein. No-go decay (NGD) detects and degrades these mRNAs. (D) Transcripts without stop codons are translated until the ribosome reaches the 3′ end of the mRNA. The potential production of elongated protein is prevented by degradation of the mRNA by non-stop decay (NSD). Alternatively, translation of the poly(A) tail leads to stalling of the ribosome and induction of NGD.

1.3 Mechanism of eukaryotic translation termination

Since NMD has the potential to discriminate between normal and abnormal termination codons, it is important for the understanding of NMD to analyze the molecular events occurring during translation termination. During the elongation phase of translation, the eukaryotic elongation factor 1 (eEF1) complex guides cognate aminoacyl-tRNA to the A site of the ribosome in order to elongate the peptide chain (Figure 3, step 1) (Sasikumar et al., 2012). However, when the ribosome encounters a stop codon (UAA, UGA or UAG) in the A site, this codon is not recognized by tRNA, but by the eukaryotic release factors 1 and 3 (eRF1 and eRF3) (Jackson et al., 2012; Klaholz, 2011).

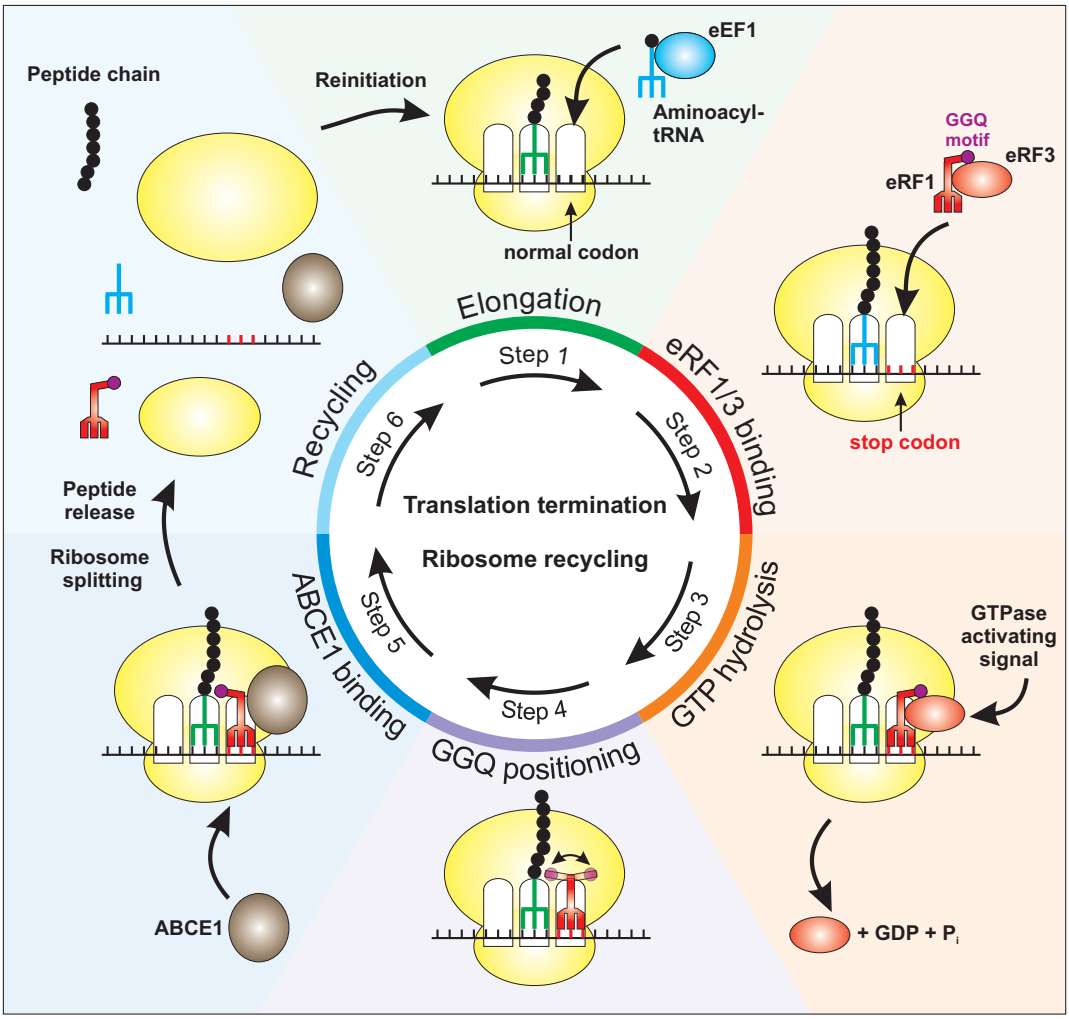


Figure 3: Schematic overview of eukaryotic translation termination. After initiation of translation, aminoacyl-tRNAs are recruited to the ribosome by eEF1 in the elongation phase in order to generate the peptide chain (Step 1). Upon entry of a stop codon in the A site, eRF1 and eRF3 interact with the ribosome and decode the stop codon (Step 2). Hydrolysis of the eRF3-bound GTP can be activated by interacting factors and results in the dissociation of eRF3 from the ribosome (Step 3). Thereby, the GGQ motif of eRF1 can be positioned properly to enable the hydrolysis of the tRNA-peptide bond (Step4). This positioning is further enhanced by the association of the recycling factor ABCE1, which fills the space previously occupied by eRF3 (Step 5). Translation termination is completed by ATP-mediated splitting of the ribosomal subunits, accompanied by peptide hydrolysis mediated by eRF1 (Step 6). Of note, peptide release can also take place during steps 4 and 5. Finally, the single components are used for another round of translation.

The proteins eRF1 and eRF3 form a complex which is structurally reminiscent to the tRNA-eEF1 complex, with eRF1 decoding the stop codon via multiple conserved sequence motifs and eRF3 serving as the eRF1-delivering factor (Figure 3, step 2) (Kong et al., 2004; Song et al., 2000). Furthermore, eRF1 catalyzes the hydrolysis of the peptidyl-tRNA ester bond using a GGQ motif that can be positioned in the peptidyl transferase center of the ribosome (Cheng et al., 2009). To accomplish this step, the GTPase eRF3 has to hydrolyze GTP and dissociate from the ribosome, therefore making space for the main recycling factor ABCE1 (Figure 3, step 3) (Pisarev et al., 2010; Salas-Marco and Bedwell, 2004). Either because of the GTP hydrolysis and dissociation of eRF3 or because of the association of the ATPase ABCE1 with the ribosome,

conformational changes of eRF1 are induced, therefore correctly positioning the GGQ motif for hydrolysis (Figure 3, steps 4 and 5) (Becker et al., 2012; Franckenberg et al., 2012). The final step is the ATP-hydrolysis induced splitting of the ribosome followed by the recycling of the ribosomal subunits for another round of translation (Figure 3, step 6). At some point during these steps of translation termination, the decision whether the stop codon is considered normal or aberrant has to be made. Due to its central role in regulating the progression in the translation termination pathway, eRF3 is considered to be involved in this decision-making process (Franckenberg et al., 2012). This is further supported by structural data obtained by cryo-electron microscopy (cryo-EM), suggesting that the flexible N-terminus and the GTPase domain of eRF3 are positioned outside of the ribosome and are solvent-exposed (Preis et al., 2014; Taylor et al., 2012). Thereby, these domains are likely available for binding of potential GTPase-modulating factors, which influence the further advancement in translation termination (Figure 3, step 3). It is therefore conceivable that, depending on the type of the eRF3 interaction partner, either the current termination event proceeds normally and without mRNA degradation or the mRNA is marked as aberrant and is subsequently degraded.

1.4 Models of NMD activation and substrate definition

As discussed above, certain factors or elements need to exist on the mRNP, which initiate NMD during translation termination. Despite NMD being a general and evolutionary conserved mRNA surveillance mechanism, different models for the activation in various organisms have been proposed (Rebbapragada and Lykke-Andersen, 2009; Schweingruber et al., 2013). Nevertheless, the key underlying determinant for NMD-induction is similar: translation is terminated at an unusual or aberrant position on the mRNP. In lower eukaryotes, the distance of the stop codon to the poly(A) tail at the 3′ end of the mRNA is a critical determinant for the recognition of PTCs (Amrani et al., 2004; Muhlrad and Parker, 1999). When a PTC is introduced in the transcript, the resulting elongated 3′ UTR is believed to disturb interactions between the terminating ribosome and downstream factors, which are required for proper termination (Figure 4A; also see 1.3). The poly(A) binding protein (PABP, in yeast Pab1 and in mammals PABPC1), which binds the poly(A) tail of mRNAs via its two first RNA recognition motifs (RRM), is one of these potential downstream factors (Adam et al., 1986; Deo et al., 1999). Specifically, the direct interaction of eRF3 and PABP is believed to be impaired due to the long distance between stop codon and poly(A) tail (Figure 4B). Interestingly, this interaction is differently mediated in yeast and mammals, as the mammalian PAPBC1 interacts via its C-terminal MLLE

domain with the N-terminal PAM2 motifs of eRF3, whereas the yeast bindings sites are less well defined (Cosson et al., 2002; Kozlov and Gehring, 2010; Kozlov et al., 2001; Roque et al., 2015). Since PABP binding stimulates the GTP hydrolysis of eRF3, loss of this interaction results in decreased translation termination efficiency (Amrani et al., 2006; Hoshino et al., 1999; Kononenko et al., 2010; Uchida et al., 2002). This in turn is supposed to enable proteins from the NMD machinery to interact with eRF3 in order to activate NMD (Czaplinski et al., 1998; Ivanov et al., 2008; Kashima et al., 2006; Singh et al., 2008; Wang et al., 2001). This "faux 3′ UTR" model of NMD activation was further supported by the observation that artificial recruitment of PABP closely downstream of a PTC suppresses NMD (Amrani et al., 2004; Behm-Ansmant et al., 2007a; Silva et al., 2008). Despite many observations being in agreement with the faux 3′ UTR model, recent studies revealed discrepancies, which are not covered by this model (Kervestin et al., 2012; Meaux et al., 2008; Roque et al., 2015).

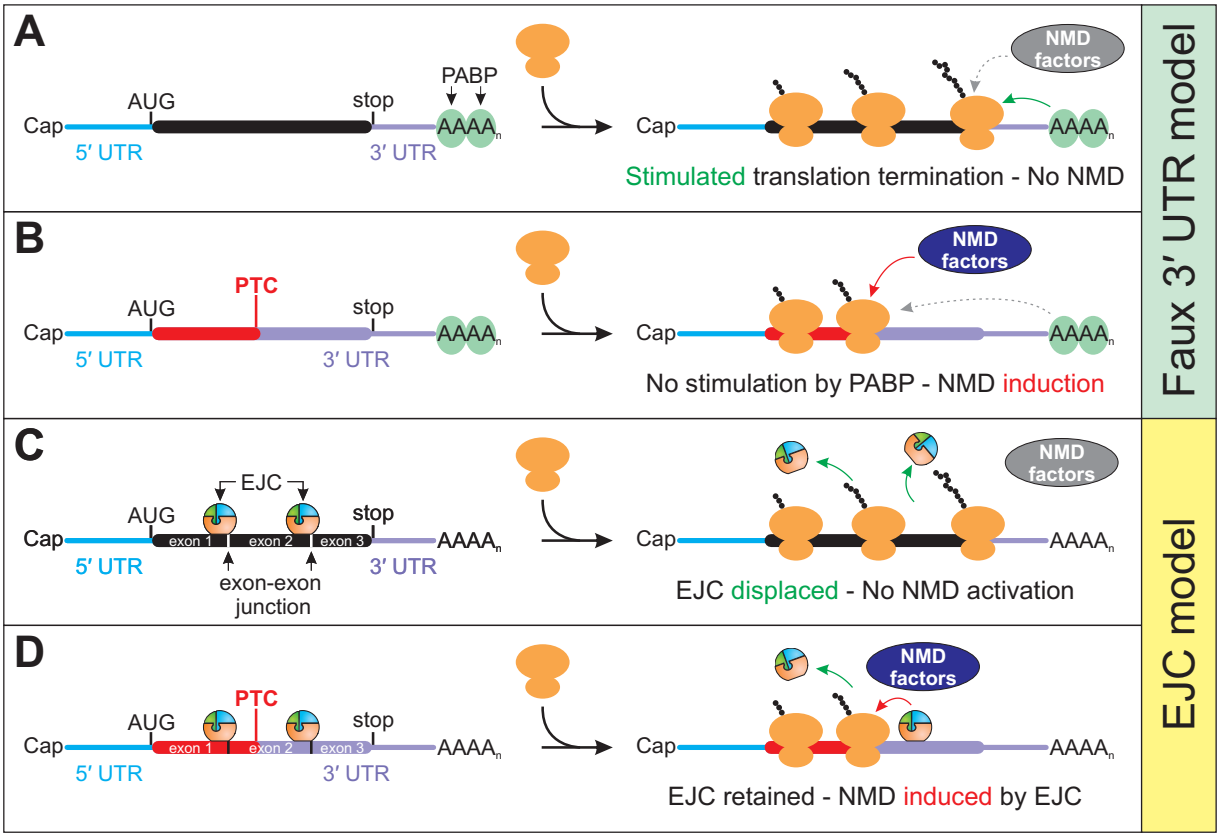


Figure 4: Key concepts of NMD activation. (A) In transcripts with short 3′ UTR, the stop codon and the downstream poly(A) tail populated by cytoplasmic poly(A) binding proteins (PABP) are positioned in close proximity. Translation termination is stimulated by PABP and therefore prevents NMD factors to initiate degradation. (B) According to the faux 3′ UTR model, translation of PTC-containing mRNA results in ribosome stalling at a position where PABP is unable to efficiently promote translation termination. In turn, NMD factors can initiate the NMD pathway. (C) EJCs deposited on the mRNA are displaced and removed by the translating ribosome, preventing the EJC-induced recruitment of NMD factors. (D) The presence of a PTC sufficiently upstream of an exon-exon junction results in incomplete EJC removal. This enables the EJC to initiate NMD.

Contrary to the situation in yeast, early studies showed that PTCs are distinguished from normal stop codons in mammalian cells when a EJC is located downstream of the terminating ribosome

(Sun et al., 2000; Thermann et al., 1998; Zhang et al., 1998a; Zhang et al., 1998b). This is in accordance with the observation that normal stop codons are positioned either in the last exon or are followed by introns not more than 50 nucleotides downstream (Brocke et al., 2002; Maquat and Li, 2001; Nagy and Maquat, 1998). As a consequence of this so-called position rule, PTCs introduced in the last exon will escape detection, whereas PTCs in any other exon will elicit NMD, if not positioned too close to the last splice site. The explanation for this rule is that during translation all EJCs in the ORF are removed by the ribosome, whereas EJCs in the 3′ UTR are unaffected and their position is maintained (Figure 4C and D) (Dostie and Dreyfuss, 2002; Lejeune et al., 2002). The required minimum distance between stop codon and EJC is necessary, because closer positioning would already result in EJC dismantling due to steric reasons. According to the current model, downstream EJCs recruit NMD factors and thereby define an upstream stop codon as premature (Figure 4D).

It was initially believed that the EJC-induced NMD activation can only occur on mRNAs which are translated for the very first time (Chiu et al., 2004; Ishigaki et al., 2001; Matsuda et al., 2007; Sato et al., 2008). This so-called “pioneer round of translation” is the first loading of ribosomes on the mRNA and is characterized by the nuclear cap-binding complex (CBC) heterodimer, consisting of CBP80 and CBP20, still attached to the 7-methylguanosine cap (Maquat et al., 2010). After export of the mRNP from the nucleus, the CBC is replaced by the cytoplasmic eIF4F complex, consisting of eIF4A, eIF4E and eIF4G (Figure 1A) (Gross et al., 2003). Recent studies showed that EJC-induced NMD takes place also on mRNPs bound by eIF4F, therefore NMD is not limited to the very first translation event (Durand and Lykke-Andersen, 2013; Rufener and Muhlemann, 2013).

Several studies reported that EJC-independent NMD exists in mammalian cells as well, which exhibits features comparable to the yeast “faux 3′ UTR” model (Buhler et al., 2006; Eberle et al., 2008; Singh et al., 2008). Substrates for this NMD pathway are normally devoid of EJCs downstream of the stop codon but contain an elongated 3′ UTR. Reporter mRNAs with artificially inserted, unspliced regions of various lengths (600-1700 nucleotides) in the 3′ UTR resulted in decreased reporter levels and accelerated degradation via NMD (Buhler et al., 2006; Eberle et al., 2008; Huang et al., 2011; Singh et al., 2008). Interestingly, endogenous transcripts with long 3′ UTRs, which encode for full length protein, are also targeted for degradation. Therefore, NMD is not restricted to degrade faulty mRNAs that arise due to errors during gene

expression, but it also regulates wild type transcripts. Accordingly, genome-wide analyses in various eukaryotic organisms showed that about 3-10% of all cellular mRNAs are upregulated upon NMD inhibition (Guan et al., 2006; He et al., 2003; Johansson et al., 2007; Lelivelt and Culbertson, 1999; Mendell et al., 2004; Ramani et al., 2009; Rehwinkel et al., 2005; Tani et al., 2012; Wittmann et al., 2006; Yepiskoposyan et al., 2011).

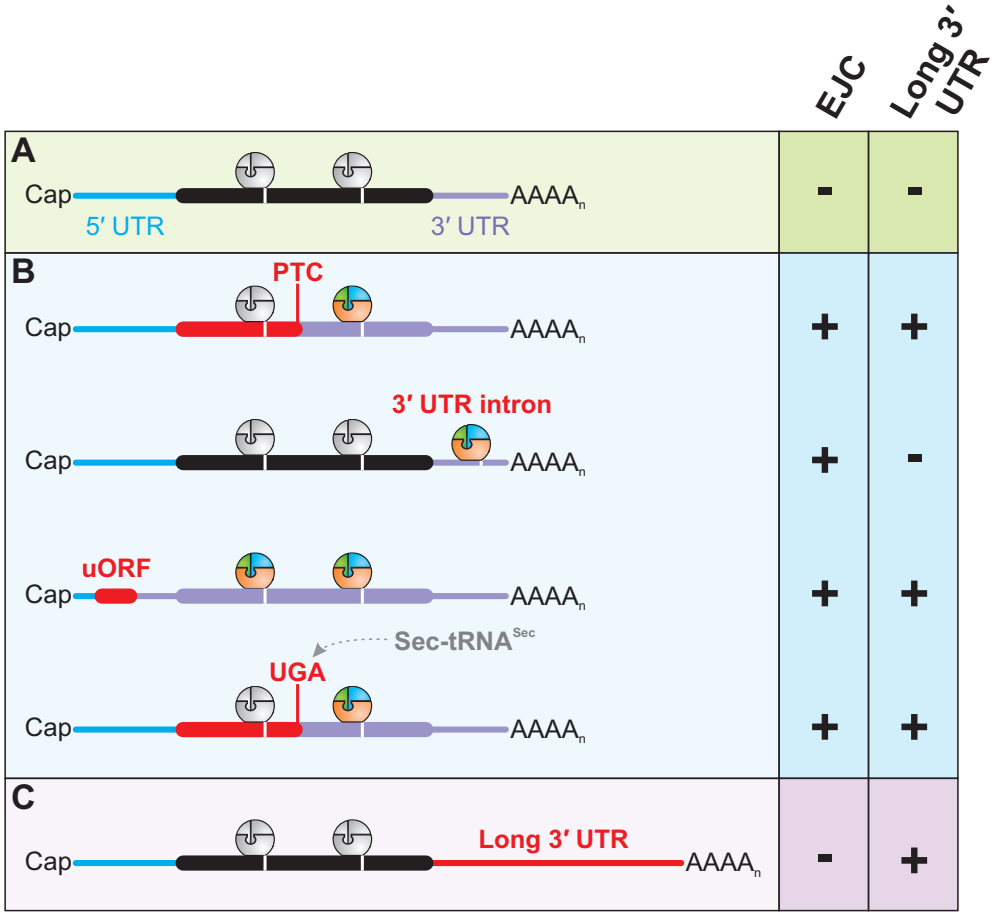


Figure 5: Examples of mRNPs targeted by NMD. (A) Normal mRNP composition with no EJC or long 3′ UTR downstream of the stop codon represents a NMD-resistant transcript. (B) Different causes exist for irregular EJC positioning relative to the stop codon, leading to NMD. Mutations, errors during transcription or splicing, and induced frameshifts are examples for generating a PTC in the regular ORF of the mRNP (top). Regulated splicing of 3′ UTR introns results in the deposition of EJC downstream of the physiological stop codon (second from top). Usage of upstream ORF (uORF) for translation initiation normally leads to premature translation termination, rendering all downstream EJC as potentially NMD-active (second from bottom). Lack of aminoacyl-tRNA^{Sec} which would encode the stop codon UGA for selenocysteine, results in premature translation termination. (C) Elongated distance between stop codon and poly(A) tail activates NMD.

In the light of the several molecular circumstances, which can potentially lead to NMD activation, NMD targets represent a heterogeneous class of mRNPs (Figure 5). NMD-resistant mRNPs in general contain the correct 3′ UTR architecture required for proper translation termination and lack downstream EJCs (Figure 5A). In contrast, various reasons for EJC-induced degradation of mRNPs exist (Figure 5B). As already discussed, PTC-containing transcripts, which can be generated by a multitude of potential errors during gene expression, frequently exhibit EJCs downstream of the PTC. In specific cases, introns are positioned in the regular 3′ UTR and

splicing of these introns generates a PTC-like situation. One example for such NMD target is the mRNA encoding for the serine/arginine-rich (SR) splicing factor SC35 (also referred to as SRSF2). SC35 regulates alternative splicing of its own mRNA in a concentration-dependent manner (Sureau et al., 2001). High SC35 levels lead to the excision of a 3′ UTR-located intron, resulting in degradation of the mRNA via EJC-induced NMD. SC35 therefore utilizes the NMD pathway for autoregulatory purposes by specifically activating NMD when required. A different class of NMD targets is degraded not because the ORF or 3′ UTR is modified, but because an upstream ORF (uORF) located in the 5′ UTR is translated. Since translation initiated at uORFs normally terminates upstream of the original ORF, not only the 3′ UTR is massively elongated, but also all normally displaced EJCs are still present. As long as no reinitiation of the ribosome occurs further downstream, this transcript will be degraded by NMD (Neu-Yilik et al., 2011). One class of NMD targets encode for selenoproteins, which are characterized by the incorporation of selenocysteine (Sec) by the UGA codon. Upon low selenium levels in the cell, the $tRNA^{Sec}$ cannot be aminoacylated and the UGA codon will be recognized as a stop codon (Moriarty et al., 1998). This can in turn lead to NMD activation, given that the UGA codon location results in an elongated and/or EJC-populated 3′ UTR. EJC-independent NMD targets, as mentioned earlier, are normally degraded because of their unusually long 3′ UTR (Figure 5C).

1.5 Factors involved in NMD assembly

Once a termination codon has been identified as aberrant, the NMD machinery has to properly assemble to execute the degradation of the target. Understanding of this process requires the detailed knowledge of the involved proteins and their molecular functions. The first proteins critical for NMD were discovered in nonsense suppression screens performed in *S. cerevisiae* and *C. elegans* (Culbertson et al., 1980; Hodgkin et al., 1989). The identified yeast *upf* (up-frameshift) and worm *smg* (suppressor with morphogenetic effect on genitalia) mutations were characterized later and the responsible genes were termed UPF1-3 (Cui et al., 1995; Leeds et al., 1991; Leeds et al., 1992) and SMG1-7 (Cali et al., 1999; Hodgkin et al., 1989; Pulak and Anderson, 1993), respectively. UPF1-3 are the evolutionary central core of the NMD factors, as homologs have been identified in all late-branching eukaryotes (Behm-Ansmant et al., 2007b; Chen et al., 2008; Culbertson and Leeds, 2003; Kadlec et al., 2006). The SMG proteins seem to have evolved later and are found, with exceptions, mostly in metazoans. The initially characterized *C. elegans* SMG2-4 proteins are homologous to the yeast UPF1-3 proteins, therefore the extended mammalian NMD core factors consist of UPF1-3, SMG1, and SMG5-7

(Applequist et al., 1997; Aronoff et al., 2001; Denning et al., 2001; Lykke-Andersen et al., 2000; Ohnishi et al., 2003; Page et al., 1999; Yamashita et al., 2001). To date, the number of proteins involved in NMD has doubled, although for many the specific role in NMD has not been characterized in detail. In mammalian cells, these include the proteins SMG8, SMG9, PNRC2, DHX34, NBAS, RUVBL1, RUVBL2, MOV10, GNL2 and SEC13 (Casadio et al., 2015; Gregersen et al., 2014; Hug and Caceres, 2014; Izumi et al., 2012; Longman et al., 2013; Longman et al., 2007; Yamashita et al., 2009).

1.5.1 The RNA helicase UPF1 plays a central role in NMD

Research on NMD has so far been consistent in the point that the evolutionary highly conserved UPF1 is the most essential NMD factor in all investigated organisms. This is because UPF1 represents the center of the NMD machinery as it interacts with a multitude of other core factors and is functionally involved in all stages from the initiation until the disassembly of the NMD complex.

Early studies proposed that the release factors eRF1 and eRF3 directly recruit UPF1 to mRNA targets in order to initiate the NMD pathway (Kashima et al., 2006). This would imply that UPF1 is loaded onto the transcript in a translation-dependent and regulated manner. However, individual-nucleotide-resolution UV cross-linking and immunoprecipitation (iCLIP) experiments showed that UPF1 has the ability to bind mRNAs even in the absence of active translation (Zund et al., 2013). Furthermore, UPF1 binds NMD targets and those that are NMD-resistant to an equal extent, suggesting that a regulated loading on NMD targets is unlikely. Furthermore, UPF1 occupies preferentially the 3′ UTR region of mRNA due to displacement from the 5′ UTR and coding region by scanning and translating ribosomes, respectively (Hurt et al., 2013; Kurosaki and Maquat, 2013; Zund et al., 2013). Since UPF1 is able to compete with PABPC1 for binding to eRF3, it was proposed that long 3′ UTRs as NMD-activating elements not only increase the distance between PABPC1 and eRF3, but also increase the local concentration of the competitor UPF1 (Hogg and Goff, 2010; Singh et al., 2008; Zund et al., 2013). The molecular details of this mechanism, specifically how the eRF3-UPF1 interaction could initiate the NMD pathway, is still unclear.

Concerning the domain architecture, the central part of UPF1 comprises two functional domains, the N-terminal zinc knuckle cystidine-histidine-rich CH domain followed by the central helicase domain formed by two RecA-like domains (Figure 6A) (Culbertson and Leeds, 2003).

Because of sequence motif composition and the ATP-hydrolysis driven ability of UPF1 to unwind nucleic acid duplexes in the 5′-3′ direction *in vitro*, the helicase belongs to the superfamily 1Bα (SF1Bα) (Bhattacharya et al., 2000; Cheng et al., 2007; Fairman-Williams et al., 2010; Singleton et al., 2007). Besides conferring potential unwinding ability, the helicase domain also mediates the direct binding to RNA (Bhattacharya et al., 2000; Chamieh et al., 2008). The overall importance of a functional UPF1 helicase domain is represented by the fact that the ATPase activity and direct RNA binding ability are both required for NMD (Mendell et al., 2002; Weng et al., 1996a, b). It remains controversial whether UPF1 utilizes ATP hydrolysis to translocate on the mRNA or uses it to remodel the mRNP after NMD execution is finished. In the first scenario, it was proposed that the helicase activity could help to bridge the distance between a terminating ribosome and the downstream-located EJC (Shigeoka et al., 2012). In the latter, it would help to recycle NMD factors and allow the execution of full exonucleolytic degradation of the mRNA once initial decay steps have taken place (Cheng et al., 2007; Franks et al., 2010; Singleton et al., 2007).

It was shown that both the CH domain as well as a C-terminal region of UPF1 (regulatory SQ region, RSQ) can regulate the helicase activity, which ensures that UPF1 clamps to the RNA and does not translocate during the earlier stages of NMD (Chakrabarti et al., 2011; Fiorini et al., 2013). Therefore, usage of the helicase domain for enabling a direct interaction with downstream factors on the mRNP in the activation phase of NMD seems unlikely. More specifically, conformational changes induced by the direct interaction of the UPF1 CH domain with the RecA2 domain results in tighter RNA binding, which represses the helicase activity (Chakrabarti et al., 2011). In order to initiate the unwinding activity of UPF1, the CH domain has to be removed from the helicase core, which is achieved by the interaction with the C-terminal UPF1-binding domain (U1BD) of UPF2 (Figure 6B and C) (Chakrabarti et al., 2011; Chamieh et al., 2008).

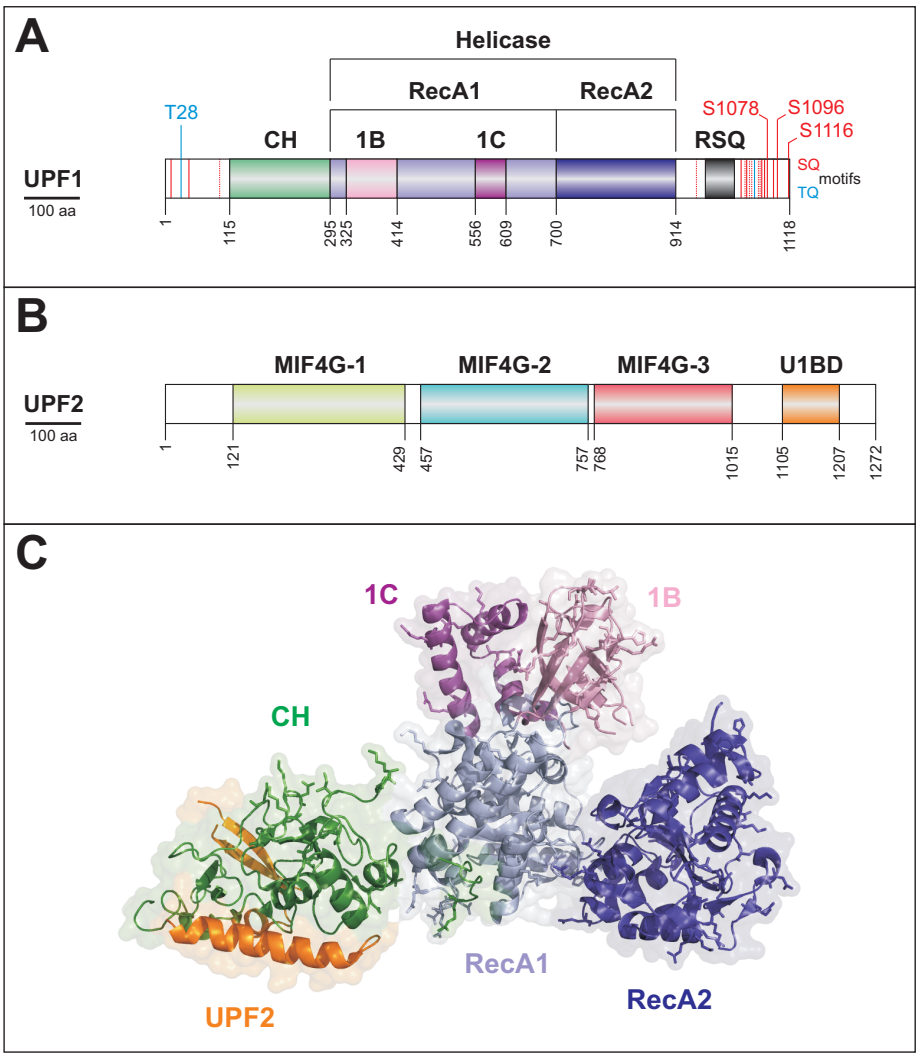


Figure 6: Interaction and domain architecture of UPF1 and UPF2. (A) Schematic domain representation of UPF1, indicating functional domains. The SQ and TQ motifs, which are potentially phosphorylated in the N- and C-terminus are indicated, the major functional ones are highlighted. Two insertions in the RecA1 domain, called 1B and 1C are unique for UPF1. (B) UPF2 domains are depicted as in (A), highlighting the three MIF4G domains and the UPF1-binding domain (U1BD). (C) Crystal structure of UPF1 core domains and co-crystallized UPF2 C-terminus. Atomic coordinates of PDB accession code 2WJV were modeled with PyMol (Schrodinger, 2010) according to Clerici et al. (2009).

1.5.2 UPF2 provides the scaffold for the NMD assembly

As a core factor for NMD, UPF2 has additional roles besides the above-discussed stimulation of UPF1 helicase activity. UPF2 consists of three tandem MIF4G domains (Middle portion of eIF4G), followed by the U1BD (Figure 6B) (Aravind and Koonin, 2000; Clerici et al., 2014; Ponting, 2000). MIF4G domains frequently provide the surface for critical interactions for factors involved in general mRNP metabolism (Ponting, 2000). In line with this role, the MIF4G-3 domain of UPF2 interacts with UPF3, establishing a physical bridge between UPF1 and UPF3 (Chamieh et al., 2008; Kadlec et al., 2004; Serin et al., 2001). Cryo-EM studies identified that the three N-terminal MIF4G domains form a ring-like structure together with the C-terminal U1BD (Melero et al., 2012). Besides providing potential structural functions, the role of the two N-terminal MIF4G domains in mammalian NMD is unclear. In *S. cerevisiae*, conserved residues

on the surface of the N-terminal helices of MIF4G-1 were shown to be essential for NMD, and although potential interaction partners were identified, the function of these interactions in the molecular pathway of NMD remain uncertain (Fourati et al., 2014). Although UPF2 is widely accepted as an essential NMD component in mammalian cells, UPF2-independent NMD has been observed in tethering assays (Gehring et al., 2005).

1.5.3 UPF3 acts as the link between UPF proteins and the EJC

Whereas in yeast and other invertebrates only one UPF3 protein exists, higher eukaryotes contain two UPF3 paralogs with high sequence similarity, UPF3a and UPF3b, the latter being expressed from the X chromosome in mammals (Lykke-Andersen et al., 2000; Serin et al., 2001). UPF3b was found to be the predominant NMD factor of both paralogs. However, a cross-regulatory circuit was described, which mainly involves the regulation of UPF3a stability as a consequence of the competition of both UPF3 proteins for binding to UPF2 (Chan et al., 2009; Gehring et al., 2003; Kunz et al., 2006). UPF3b is a nucleocytoplasmic shuttling protein and contains a conserved N-terminal RNA recognition motif (RRM). This domain is the binding site for the MIF4G-3 of UPF2 and does not mediate RNA binding (Kadlec et al., 2004; Lykke-Andersen et al., 2000; Serin et al., 2001). At the C-terminus, a short linear motif termed EJC-binding motif (EBM) is responsible for the interaction of UPF3b with a composite binding site of the EJC formed by the core components eIF4A3, MAGOH and Y14 (Buchwald et al., 2010; Chamieh et al., 2008; Gehring et al., 2003; Kashima et al., 2010). UPF3b likely associates with the EJC in the nucleus and remains bound until it is displaced by PYM during ribosome-mediated EJC disassembly (Bono and Gehring, 2011; Chamieh et al., 2008; Gehring et al., 2003; Tange et al., 2004). It was proposed that for mammalian NMD, EJCs downstream of a translation termination event could increase the concentration of UPF1-UPF2 in the mRNP due to the specific recruitment via UPF3b (Kervestin and Jacobson, 2012). Yet, the exact molecular function of UPF3 in NMD remains elusive, since this UPF1-UPF2 recruiting function does not explain the function of UPF3 in EJC-independent NMD (Chamieh et al., 2008; Melero et al., 2012; Metze et al., 2013). This is especially interesting in case of organisms that do not employ EJC-enhanced NMD as the standard pathway, but still rely on UPF3 for NMD. Examples are yeast, flies and worms, which either contain a very small number of spliced transcripts, lack EJC proteins and the EBM in the C-terminus of UPF3, or do not require EJC core components for NMD, respectively (Culbertson and Leeds, 2003; Gatfield et al., 2003; Gehring et al., 2003; Longman et al., 2007; Spingola et al., 1999; Wen and Brogna, 2010). Similar to the UPF2-

independent NMD described, UPF3-independent pathways were observed also, suggesting that these proteins are not absolutely necessary for certain NMD events (Chan et al., 2007).

1.5.4 UPF1 is phosphorylated by the SMG1 kinase

It was first observed in *C. elegans* that UPF1 (called SMG2 in *C. elegans*) is a phosphoprotein (Page et al., 1999). The phosphorylation status of UPF1 was found to be positively regulated by SMG1, UPF2 and UPF3 (SMG3 and SMG4 in *C. elegans*) and negatively by SMG5-7. SMG1 belongs to the phosphatidylinositol (PI) 3-kinase-related kinase (PIKK) family and was characterized as the responsible kinase for UPF1 phosphorylation in metazoan cells (Denning et al., 2001; Grimson et al., 2004; Page et al., 1999; Yamashita et al., 2001). Structural studies showed that the domain arrangement of the 410 kDa SMG1 protein is divided into a catalytic head structure and a flexible arm (Figure 7A) (Arias-Palomo et al., 2011; Melero et al., 2014). The binding of regulatory proteins termed SMG8 and SMG9 to the arm region of SMG1 modulates the kinase activity of SMG1 (Arias-Palomo et al., 2011; Fernandez et al., 2011; Yamashita et al., 2009). The head region including the catalytic PIKK domain and the FRB domain mediates the interaction with UPF1 and UPF2, respectively (Melero et al., 2014). The UPF2 binding to SMG1 is believed to modulate and positively stimulate the kinase activity, resulting in the phosphorylation of UPF1 (Ivanov et al., 2008; Kashima et al., 2006). Thereby, the list of potentially essential roles for UPF2 can be extended, since UPF2 not only forms the linear interaction cascade from UPF1 to UPF3 and modulates the helicase of UPF1, but it also positively influences the phosphorylation of UPF1.

1.5.5 Initiation of mRNA degradation via phospho-UPF1 interactions

PIKK members, like SMG1, preferentially phosphorylate serines and threonines followed by glutamines (SQ and TQ motifs) (Bensimon et al., 2011; Yamashita et al., 2001). The SMG1-phoshorylated SQ and TQ motifs of mammalian UPF1 are clustered in the extended and unstructured N- and C-terminus (Figure 6A) (Chakrabarti et al., 2014; Page et al., 1999; Yamashita et al., 2001). Even though phosphorylation was also reported for yeast Upf1, the mechanism and responsible kinase are different, since yeast Upf1 lacks most of the clustered SQ and TQ motifs in the C-terminus and no orthologue of SMG1 has been found (Lasalde et al., 2014; Wang et al., 2006). The phosphorylation sites in mammalian UPF1 act as recruitment platforms for the remaining core NMD factors, SMG5, SMG6 and SMG7. The three proteins

share one common domain feature, a 14-3-3-like domain which folds similar to 14-3-3 proteins and is able to interact with phosphorylated peptides (Figure 7B) (Fukuhara et al., 2005).

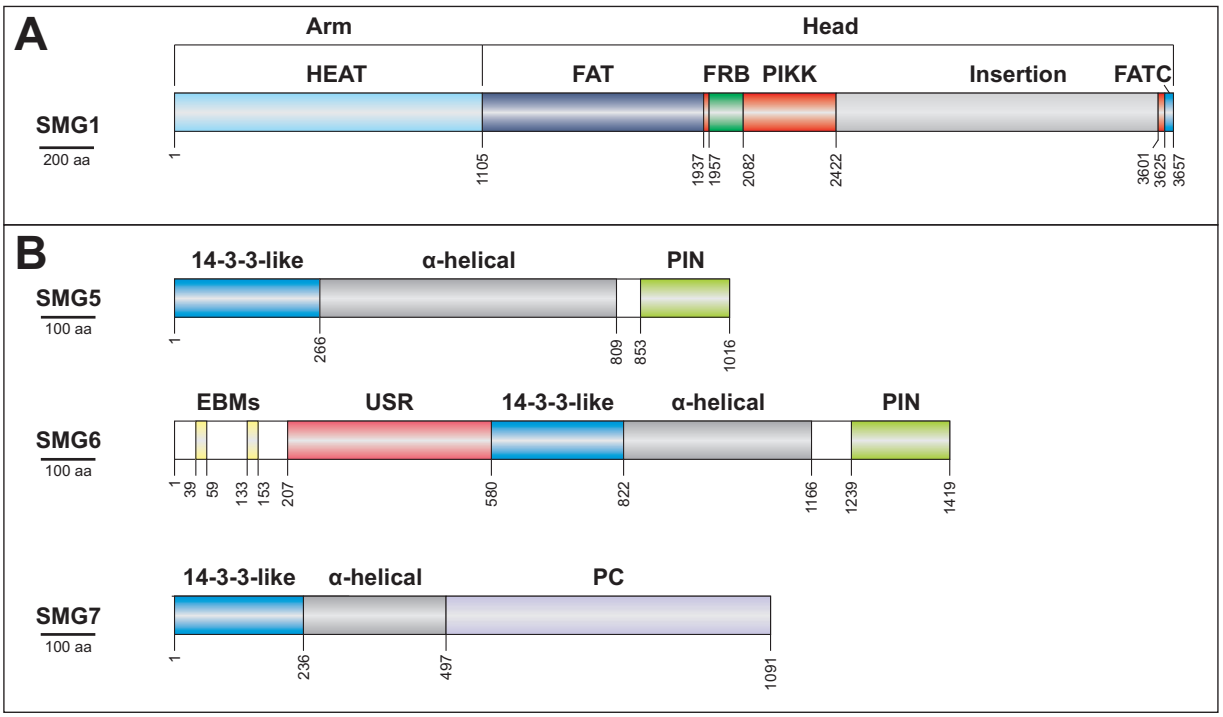


Figure 7: Domain structure of SMG proteins. (A) The complex domain architecture of SMG1 is depicted schematically. The N-terminal HEAT repeats form the arm, whereas the C-terminal domains form the globular head. (B) The decay inducing SMG5-7 proteins share a 14-3-3-like domain, which is followed by α-helical extensions required for the stabilization of the domain. The remaining domains or functional regions are indicated. PC = proline-rich region.

SMG5 and SMG7 form a heterodimer by perpendicular back-to-back interactions of their N-terminal 14-3-3-like domains. This is an uncommon arrangement, compared to the normal head-to-head interaction found in most 14-3-3 dimers and could explain why normal 14-3-3 proteins do not interact with UPF1 (Gardino et al., 2006; Jonas et al., 2013; Obsil and Obsilova, 2011). The 14-3-3-like domain of SMG7 is mostly responsible for the phosphorylation-dependent interaction between phosphorylated amino acids (e.g. S1096) in the C-terminus of UPF1 and the heterodimer SMG5-SMG7 (Chakrabarti et al., 2014; Fukuhara et al., 2005; Jonas et al., 2013; Okada-Katsuhata et al., 2012). The 14-3-3-like domain of SMG5, which by itself is not able to interact with UPF1, supposedly provides additional binding strength and specificity (Jonas et al., 2013; Okada-Katsuhata et al., 2012).

1.5.6 Initiation of exonucleolytic degradation

Early work showed that artificial recruitment of full length SMG7 or the C-terminal proline-rich (PC) region to tethering reporter mRNA induces mRNA degradation in a position-independent and XRN1-/DCP2-dependent manner (Unterholzner and Izaurralde, 2004). DCP2 is the catalytic subunit of the decapping complex, whereas XRN1 is the major cytoplasmic 5′-3′ exonuclease,

suggesting that SMG7 induces accelerated decapping (Ghosh and Jacobson, 2010). Recently, the direct interaction of the PC region of SMG7 with the catalytic subunit of the CCR4-NOT deadenylase complex POP2 has been shown (Loh et al., 2013). Therefore, SMG7 recruitment, mediated by its 14-3-3-like domain, to the C-terminus of phospho-UPF1 induces deadenylation followed by decapping and degradation of the mRNA in the 5′-3′ direction (Loh et al., 2013). Early reports showed that the N- and C-terminus of UPF1 can interact with decapping proteins, however, it was unclear if this interaction is direct or mediated by another factor (He and Jacobson, 1995, 2001; Lejeune et al., 2003; Lykke-Andersen, 2002). The proline-rich nuclear receptor coregulatory protein 2 (PNRC2) interacts with UPF1 and the decapping complex component DCP1, thereby providing a potential link for deadenylation-independent decapping during NMD (Cho et al., 2009; Lai et al., 2012).

1.5.7 Dephosphorylation of UPF1 is initiated by decay factors

NMD is impaired under conditions where UPF1 accumulates in the hyper- or hypo-phosphorylated form, suggesting that a cycle of phosphorylation and dephosphorylation is essential (Grimson et al., 2004; Ohnishi et al., 2003; Okada-Katsuhata et al., 2012; Page et al., 1999; Yamashita et al., 2001). Protein phosphatase 2A (PP2A) associates with the SMG5-SMG7 heterodimer via the interaction with SMG5 and was identified as the phosphatase required for the dephosphorylation of UPF1 (Anders et al., 2003; Ohnishi et al., 2003). SMG5 contains a C-terminal PilT N-terminus (PIN) domain, which is potentially involved in the interaction with PP2A. Deletion of the very C-terminal amino acids or the replacement of a conserved aspartate at position 860 in this domain increased the phosphorylation of UPF1 (Ohnishi et al., 2003). PIN domains are commonly found in proteins executing endonuclease activity, however, the catalytic triad normally consisting of three aspartate residues is absent in the SMG5 PIN domain and no endocleavage activity was reported neither *in vivo* nor *in vitro* (Clissold and Ponting, 2000; Glavan et al., 2006; Schoenberg, 2011). Interestingly, D860 is the one remaining aspartate residue in the active site, which was implicated in the regulation of UPF1 phosphorylation status (Ohnishi et al., 2003). Of note, evidence for SMG6 association with the PP2 complex was provided as well, suggesting that, in line with the initial observation in *C. elegans*, all three SMG5-7 proteins mediate UPF1 dephosphorylation by recruiting phosphatases (Chiu et al., 2003).

1.5.8 Endonucleolytic cleavage is executed by SMG6

Studies on the preferred nucleolytic degradation pathway of PTC containing mRNA in *Drosophila melanogaster* S2 cells showed that the knockdown of exonucleolytic machineries employing deadenylation, decapping, 3′-5′ and 5′-3′ degradation could not stabilize reporter mRNA levels (Gatfield and Izaurralde, 2004). However, evidence for PTC-dependent endonucleolytic cleavage was found due to the accumulation of 3′ and 5′ fragments upon depletion of XRN1 and components of the 5′-3′ degrading exosome complex, respectively (Gatfield and Izaurralde, 2004). In metazoans, SMG6 was identified as the endonuclease responsible for cleavage of the NMD targets in the vicinity of the stop codon (Eberle et al., 2009; Gatfield and Izaurralde, 2004; Huntzinger et al., 2008). SMG6 contains a C-terminal PIN domain similar to SMG5. In contrast to SMG5, all catalytically important residues are present in the active site and the SMG6 PIN domain exhibits nucleolytic activity *in vitro* (Glavan et al., 2006). Mutations of any of the catalytic aspartate residues, which are required to coordinate divalent metal ions for the nucleophilic attack of H_2O on the phosphodiester bond of the RNA, renders the protein inactive and abolishes endonucleolytic degradation of NMD targets (Eberle et al., 2009; Glavan et al., 2006; Huntzinger et al., 2008; Kashima et al., 2010; Nicholson et al., 2014). Like SMG5 and SMG7, SMG6 contains a 14-3-3-like domain, which is located centrally in the protein and does not form hetero- or homodimers (Chakrabarti et al., 2014; Fukuhara et al., 2005). This domain was also suggested to bind phosphorylated UPF1 and mutational analysis showed that mutation of the residues in the phosphopeptide binding pocket abolished the interaction with UPF1 (Okada-Katsuhata et al., 2012). Similarly, alanine exchange of T28 in the N-terminus of UPF1 greatly reduces the interaction with SMG6, suggesting that the 14-3-3-like domain of SMG6 interacts with the phosphorylated N-terminus of UPF1 (Okada-Katsuhata et al., 2012). In recently reported *in vitro* experiments with phosphorylated UPF1, the phospho-dependent interaction with SMG5-SMG7 was confirmed, however, no interaction of the isolated 14-3-3-like domain with hyperphosphorylated UPF1 was observed (Chakrabarti et al., 2014). This is in line with recent data showing that phosphorylated UPF1 preferentially occupies the 3′ UTR of NMD targets in a complex with SMG5 and SMG7, but not SMG6 (Kurosaki et al., 2014). However, the unstructured region preceding the 14-3-3-like domain of SMG6 was observed to bind UPF1 in a phospho-independent manner *in vitro*, which was complemented by functional studies of SMG6 tethering and UPF1 complementation assays performed in another recent publication (Chakrabarti et al., 2014; Nicholson et al., 2014). In addition, two

EBMs were characterized in the very N-terminus of SMG6, which similarly to the EBM of UPF3b, mediate the interaction with the EJC (Kashima et al., 2010). These EBM motifs were found to be essential for NMD (Kashima et al., 2010). Given the multitude of possible interactions, the exact mechanisms by which SMG6 is recruited to the target mRNA remain elusive.

1.6 Model of the EJC-NMD mechanism

As the details of functions and interplay between NMD factors has been discussed above, the following model aims to present the most important steps from NMD initiation to mRNA degradation (Figure 8). This is exemplified for EJC-induced NMD, because the molecular events during long 3′ UTR-induced NMD are only poorly understood in mammalian cells.

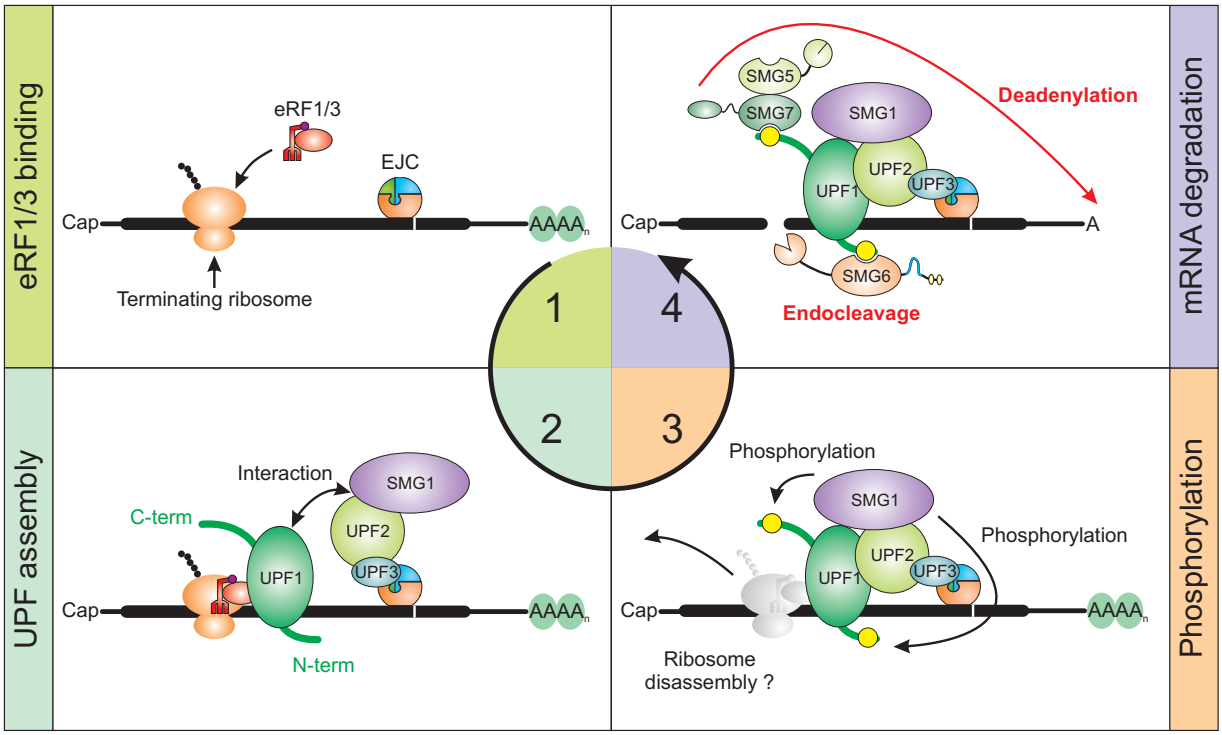


Figure 8: Schematic model of EJC-induced NMD. (1) For EJC-NMD to be initiated, the ribosome has to stall at a stop codon upstream of an EJC. (2) Following the interaction of UPF1 with eRF3, UPF2 and SMG1 are recruited. This is enhanced due to the EJC-bound UPF3 recruitment of UPF2. (3) Activated SMG1 phosphorylates UPF1 at the N- and C-terminus. It is currently unclear, if the ribosome is already disassembled at this point. (4) Phosphorylated UPF1 recruits SMG5/7 and SMG6, which initiate degradation via deadenylation or endocleavage, respectively.

In short, after association with the ribosome (Figure 8, step 1), the release factors eRF1/3 interact with mRNA bound UPF1 and by an unknown mechanism activate it to bind UPF2/SMG1 (Figure 8, step 2). This is facilitated by the UPF3-mediated recruitment of these factors. By UPF2 stimulation of the SMG1 kinase activity, UPF1 becomes phosphorylated at N- and C-terminal SQ and TQ motifs (Figure 8, step 3). These sites are recognized by SMG6 at the N-terminus, leading to the endonucleolytic cleavage of the target. Alternatively or simultaneously, SMG5/7

binds to the C-terminus of UPF1 and induces the deadenylation-dependent degradation of the mRNA (Figure 8, step 4).

1.7 Physiological function of NMD and importance in diseases

Around 30% of disease-causing mutations produce PTC-containing mRNA due to nonsense or frameshift mutations (Miller and Pearce, 2014). Moreover, 20% of all single-base pair mutations affecting the coding region and leading to diseases, produce nonsense codons (Mort et al., 2008). NMD was identified as an important modifier of the clinical outcome of these nonsense mutation-based diseases (Khajavi et al., 2006). This depends on whether the introduction of a PTC activates NMD and the mRNA is consequently degraded or the mRNA evades the NMD surveillance and a truncated protein is generated.

When PTC-mRNAs evade NMD, the produced truncated protein may be dominant-negative and could be deleterious for the cell. This is observed, for example, in the rare dominant form of β-thalassemia, which results from NMD-insensitive nonsense mutations in the β-globin gene. These transcripts produce C-terminally truncated β-globin which causes toxic precipitations (Baserga and Benz, 1988; Hall and Thein, 1994; Thein et al., 1990).

In the case of NMD-activation, monoallelic PTC mutations result in only partial depletion of the encoded protein, as the other wild type allele still supports normal translation of this gene product. These mutations generally result in a recessive pattern of inheritance, as long as the inactivation of one gene copy does not lead to haploinsufficiency (Miller and Pearce, 2014). Although NMD in general helps to prevent the synthesis of harmful or unfunctional proteins, the degradation of mRNAs, which encode truncated, but still functional proteins, is detrimental for the cell. Examples for this are mutations in the dystrophin gene, which lead to severe Duchenne muscular dystrophy (DMD) phenotypes when the mRNA is degraded, while NMD-insensitive transcripts produce functional peptides resulting in the milder Becker muscular dystrophy (BMD) (Kerr et al., 2001; Pillers et al., 1999).

Diseases caused by the depletion of functional protein via the NMD pathway are the subject of intensive research, which aims to alleviate the clinical phenotype by achieving PTC suppression (Keeling and Bedwell, 2011; Keeling et al., 2012). Current and past research focuses extensively on the modulation of read-through of stop codons. Near-cognate tRNAs which contain two of the required three nucleotides in their anticodon can in principle bind to stop codons and lead

to the incorporation of an amino acid instead of release factor-mediated translation termination. Normally, the release factors are far more potent in recognizing the stop codon, upon treatment of cells with compounds such as aminogylcosides, however, this out-competing effects is decreased. Although promising effects were observed in many different *in vitro* and *in vivo* models, the current cytotoxicity and other drawbacks of these treatments still need to be reduced in order to generate a suitable therapy (Miller and Pearce, 2014).

One alternative way for potential treatment of nonsense-associated diseases is to directly target and inhibit the NMD machinery. This requires the detailed knowledge of the molecular pathway of NMD activation and execution, in order to avoid side effects. Interestingly, the depletion of NMD factors has different phenotypical outcome depending on the complexity of the investigated organism. Whereas deletion of NMD factors in budding yeast or nematodes only results in mild phenotypes, fruit flies and zebrafish require core NMD components like UPF1 and UPF2 for embryonic development and viability (Hodgkin et al., 1989; Leeds et al., 1991; Metzstein and Krasnow, 2006; Pulak and Anderson, 1993; Rehwinkel et al., 2005; Wittkopp et al., 2009). Similarly, UPF1, UPF2 or SMG6 knockouts are embryonically lethal in mice and NMD factors are required for growth and viability of plants (Jeong et al., 2011; Li et al., 2015; Medghalchi et al., 2001; Riehs-Kearnan et al., 2012; Weischenfeldt et al., 2008; Yoine et al., 2006). If lethality is a consequence of the inactivation of NMD or because these proteins also fulfill other cellular roles is currently under debate. UPF1 has many additional functions, for example, the involvement in proper progression of the cell cycle, as depletion of UPF1, but not UPF2, results in early S-phase arrest (Azzalin and Lingner, 2006).

1.8 Aims of this work

Although the underlying principles of substrate recognition, NMD activation and mRNA degradation have been extensively studied in various organisms, we still lack the detailed understanding of fundamental processes and decisions in this pathway. In this work, several key aspects along the pathway of NMD were studied in order to refine and expand our model of the molecular mechanism of this mRNA surveillance process. Specifically the following questions were addressed:

- During translation termination, the interaction of PABPC1 with eRF3 is considered as the signal that prevents NMD. The faux 3′ UTR model predicts that because of this interaction, UPF1 is outcompeted and translation termination is accelerated. Recruitment of PABPC1 downstream of an otherwise NMD-inducing stop codon inhibits NMD. The molecular requirements for this NMD suppression are, however, not fully characterized. It remains to be investigated, whether the competition between PABPC1 and UPF1 is the only determinant deciding if NMD is activated or not. (Investigated in Fatscher et al. (2014) “The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay”)
- EJCs are deposited during splicing on the mRNA, thereby influencing downstream gene expression steps and also serving as potent NMD activating signals if positioned downstream of a termination codon. Despite knowledge of the EJC composition, its positioning on the mRNA and its functions, how and by which spliceosomal protein the EJC is loaded onto the spliced transcript is unclear. (Investigated in Steckelberg et al. (2012)“CWC22 Connects Pre-mRNA Splicing and Exon Junction Complex Assembly”)
- In the assembly of the NMD complex in mammalian EJC-NMD, UPF2 plays the role of connecting the upstream ribosome-associated UPF1 with downstream EJC-bound UPF3b. Furthermore, UPF2 regulates the helicase activity of UPF1 and is involved in the activation of SMG1-mediated phosphorylation of UPF1. Whereas specific functions were assigned to the C-terminal MIF4G3 and U1BD, it remains unsolved whether the N-terminal half of the protein, containing MIF4G1 and -2, serves an essential role during NMD as well. (Investigated in Clerici et al. (2014) “Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2”)

- EJC-induced NMD targets have been shown to be cleaved endonucleolytically close to the PTC. This endocleavage is catalyzed by SMG6, which is potentially recruited to the mRNA via different interaction routes. The mRNP requirements and NMD factors involved in triggering endocleavage are not yet defined and it is unclear if long 3′ UTR NMD targets are also eliminated using the SMG6 dependent degradation pathway. Furthermore, we lack detailed understanding of the endocleavage sites on the mRNA, as this could help to understand the mRNP composition at the point of SMG6 activation. (Investigated in Boehm et al. (2014) “3′ UTR length and mRNP composition determine endocleavage efficiencies at termination codons”)

2. Publications

The publications of this cumulative thesis are inserted in the following order:

- Fatscher, T., Boehm, V., Weiche, B., and Gehring, N.H. (2014). **The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay.** RNA 20, 1579-1592.

- Steckelberg, A.L., Boehm, V., Gromadzka, A.M., and Gehring, N.H. (2012). **CWC22 connects pre-mRNA splicing and exon junction complex assembly.** Cell reports 2, 454-461.

- Clerici, M., Deniaud, A., Boehm, V., Gehring, N.H., Schaffitzel, C., and Cusack, S. (2014). **Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2.** Nucleic acids research 42, 2673-2686.

- Boehm, V., Haberman, N., Ottens, F., Ule, J., and Gehring, N.H. (2014). **3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons**. Cell reports 9, 555-568.

The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay

TOBIAS FATSCHER,[1,2] VOLKER BOEHM,[1,2] BENJAMIN WEICHE,[1] and NIELS H. GEHRING[1]
[1]Institute for Genetics, University of Cologne, 50674 Cologne, Germany

ABSTRACT

Nonsense-mediated mRNA decay (NMD) eliminates different classes of mRNA substrates including transcripts with long 3′ UTRs. Current models of NMD suggest that the long physical distance between the poly(A) tail and the termination codon reduces the interaction between cytoplasmic poly(A)-binding protein (PABPC1) and the eukaryotic release factor 3a (eRF3a) during translation termination. In the absence of PABPC1 binding, eRF3a recruits the NMD factor UPF1 to the terminating ribosome, triggering mRNA degradation. Here, we have used the MS2 tethering system to investigate the suppression of NMD by PABPC1. We show that tethering of PABPC1 between the termination codon and a long 3′ UTR specifically inhibits NMD-mediated mRNA degradation. Contrary to the current model, tethered PABPC1 mutants unable to interact with eRF3a still efficiently suppress NMD. We find that the interaction of PABPC1 with eukaryotic initiation factor 4G (eIF4G), which mediates the circularization of mRNAs, is essential for NMD inhibition by tethered PABPC1. Furthermore, recruiting either eRF3a or eIF4G in proximity to an upstream termination codon antagonizes NMD. While tethering of an eRF3a mutant unable to interact with PABPC1 fails to suppress NMD, tethered eIF4G inhibits NMD in a PABPC1-independent manner, indicating a sequential arrangement of NMD antagonizing factors. In conclusion, our results establish a previously unrecognized link between translation termination, mRNA circularization, and NMD suppression, thereby suggesting a revised model for the activation of NMD at termination codons upstream of long 3′ UTR.

Keywords: PABPC1; eIF4G; NMD; ribosome recycling; translation termination

INTRODUCTION

NMD represents a surveillance mechanism that removes transcripts with premature translation termination codons (PTCs) from eukaryotic cells (Chang et al. 2007; Rebbapragada and Lykke-Andersen 2009; Nicholson et al. 2010). The core NMD factors are present in all eukaryotes and their activity prevents the synthesis of C-terminally truncated proteins with potentially dominant negative effects (Bhuvanagiri et al. 2010). Furthermore, NMD directly or indirectly regulates the expression of many physiological mRNAs, although only some of them contain PTCs, for example, as a result of alternative splicing or upstream open reading frames (Yepiskoposyan et al. 2011; Tani et al. 2012). The function of NMD as a general regulator of gene expression explains why NMD factors are essential for normal animal development (Hwang and Maquat 2011).

In human cells, NMD is efficiently activated when at least one intron is located >50 nt downstream from the termination codon (Thermann et al. 1998; Zhang et al. 1998). During splicing in the nucleus, exon–exon junctions are marked by exon-junction complexes (EJCs), which serve as NMD-activating signals during translation in the cytoplasm. In addition to the aforementioned EJC-dependent NMD, an alternative EJC-independent NMD pathway targets mRNAs with a long 3′ UTR (Eberle et al. 2008; Singh et al. 2008; Yepiskoposyan et al. 2011). During eukaryotic translation termination, the interaction of PABP (in humans PABPC1) with the ribosome-bound eRF3 (in humans eRF3a) stimulates polypeptide release and the subsequent recycling of ribosomes (Hoshino et al. 1999; Uchida et al. 2002). However, when the interaction of PABPC1 with eRF3a is reduced by an unusually long 3′ UTR, UPF1 binds to eRF3a and activates NMD (Singh et al. 2008). Hence, tethering of PABPC1 in the proximity of a termination codon can inhibit NMD by simulating the presence of a poly(A) tail (Amrani et al. 2004;

[2]**These authors contributed equally to this work.**
Corresponding author: ngehring@uni-koeln.de
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.044933.114. Freely available online through the *RNA* Open Access option.

Behm-Ansmant et al. 2007; Eberle et al. 2008; Silva et al. 2008; Singh et al. 2008).

In the case of premature translation termination, UPF1 is phosphorylated by SMG1 within its extended N- and C-terminal regions (Kashima et al. 2006). UPF2 binds directly to the C-terminal part of SMG1 and stimulates the phosphorylation of UPF1 (Kashima et al. 2006; Clerici et al. 2013). Phosphorylated UPF1 recruits the homologous proteins SMG5/SMG7 and SMG6, leading to the degradation of the target mRNA (Okada-Katsuhata et al. 2012). While the exonucleolytic decay is coordinated by the SMG5-SMG7 heterodimer (Loh et al. 2013), the endonucleolytic cleavage of NMD targets is mediated by the C-terminal PIN (PilT N terminus) domain of SMG6 (Glavan et al. 2006; Huntzinger et al. 2008; Eberle et al. 2009).

A large distance between the poly(A) tail and the termination codon promotes NMD. Therefore, long 3′ UTRs represent an NMD activating characteristic of endogenous NMD targets (Eberle et al. 2008; Singh et al. 2008; Yepiskoposyan et al. 2011). In view of the many endogenous mRNAs that are potentially regulated by this pathway, it is important to elucidate the molecular mechanism of NMD suppression by PABPC1. Using the MS2 tethering system, we have investigated which molecular interactions of PABPC1 are required to inhibit NMD of a reporter mRNA with a long 3′ UTR. We find that tethered PABPC1 suppresses NMD induced by a long 3′ UTR. Moreover, PABPC1 does not require the interaction with eRF3a to retain its NMD suppressing activity. In contrast, a mutant of PABPC1 unable to bind eIF4G does not inhibit NMD. Furthermore, tethered eIF4G or eRF3a suppress NMD as well. Our observations suggest a tight coupling between mRNA circularization via eIF4G and NMD suppression.

RESULTS

Reporter-bound PABPC1 increases long 3′ UTR-containing mRNA levels

PABPC1 plays a pivotal role in gene expression because it promotes mRNA circularization, facilitates ribosome recycling, and suppresses NMD at normal termination codons (Wells et al. 1998; Behm-Ansmant et al. 2007). To analyze NMD suppression by PABPC1, we used a reporter construct consisting of the triosephosphate-isomerase (TPI) open reading frame (ORF) to which we added the 3′ UTR of SMG5. The SMG5 mRNA has been previously shown to undergo NMD mediated by its long 3′ UTR (Singh et al. 2008) and owing to the presence of the SMG5 3′ UTR, the reporter mRNA is degraded (V Boehm, N Haberman, F Ottens, J Ule, and NH Gehring, in prep.). We inserted four MS2 binding sites downstream from the termination codon to enable tethering of MS2-fusion proteins to a position at the beginning of the 3′ UTR (TPI-4MS2-SMG5) (Fig. 1A). Upon coexpression (i.e., tethering) of MS2V5-tagged PABPC1, the levels of the reporter mRNA increased by a factor of four compared with the MS2V5-GST that served as negative control (Fig. 1B). This suggests that PABPC1 counteracts NMD of the reporter mRNA. To confirm that the observed increase in mRNA abundance is an NMD-specific effect and not due to general mRNA stabilization by PABPC1, we used two additional TPI control reporter constructs. In one construct, the termination codon was shifted downstream from the MS2 binding sites by deleting the original termination codon (TPI-Δter-4MS2-SMG5) (Fig. 1A). In the other construct, the four MS2-binding sites were moved to a position at the 3′ end of the 3′ UTR (TPI-SMG5-4MS2) (Fig. 1A). In both cases, tethering of PABPC1 only marginally changes the levels of the reporter mRNAs (Fig. 1C,D). To exclude possible *trans*-effects of PABPC1 expression, we coexpressed either FLAG-PABPC1 together with the TPI-4MS2-SMG5 reporter (Fig. 1E), or MS2V5-PABPC1 with a TPI-SMG5 reporter lacking MS2-binding sites (Fig. 1A,F). In both cases, we observed only slight increases in mRNA levels when PABPC1 was not directly bound. In summary, our results suggest that PABPC1 is able to antagonize NMD induced by the long 3′ UTR of the reporter mRNA when it is tethered in close proximity downstream from the termination codon.

PABPC1 stabilizes mRNA by inhibition of NMD

The previous experiments analyzed the steady-state levels of reporter mRNA. To confirm that tethered PABPC1 stabilizes the mRNA, we next determined the decay rates of the reporter mRNA upon transcriptional shutoff by Actinomycin D treatment. We observed a half-life of ∼4.8 h of the TPI-4MS2-SMG5 reporter construct when MS2V5-GST is tethered as a control (Fig. 2A). Upon PABPC1 tethering to the same reporter the mRNA is stabilized with a half-life of ∼30 h (Fig. 2A). This demonstrates that MS2V5-PABPC1 is in fact able to suppress NMD by preventing the degradation of the reporter mRNA construct. Because NMD is restricted to actively translated mRNAs (Thermann et al. 1998), we aimed to confirm that the observed effects are not caused by decreased translation rates of the reporter mRNAs. To this end, we used an N-terminally FLAG-tagged TPI-4MS2-SMG5 reporter construct, enabling us to measure translation efficiency by Western blotting and to correlate these effects with the mRNA levels detected by Northern blotting. FLAG-tagged emGFP was cotransfected as a loading control for both the mRNA as well as protein expression levels. Tethering MS2V5-PABPC1 to the FLAG-tagged TPI reporter mRNA led to a similar increase in FLAG-TPI protein and mRNA levels compared with the GST control (Fig. 2B, left), indicating that tethered PABPC1 does not change overall translation rates (Fig. 2B, right). We obtained similar results using a dual luciferase reporter system (data not shown). These results indicate that the NMD antagonizing effect of tethered PABPC1 is not caused by decreased

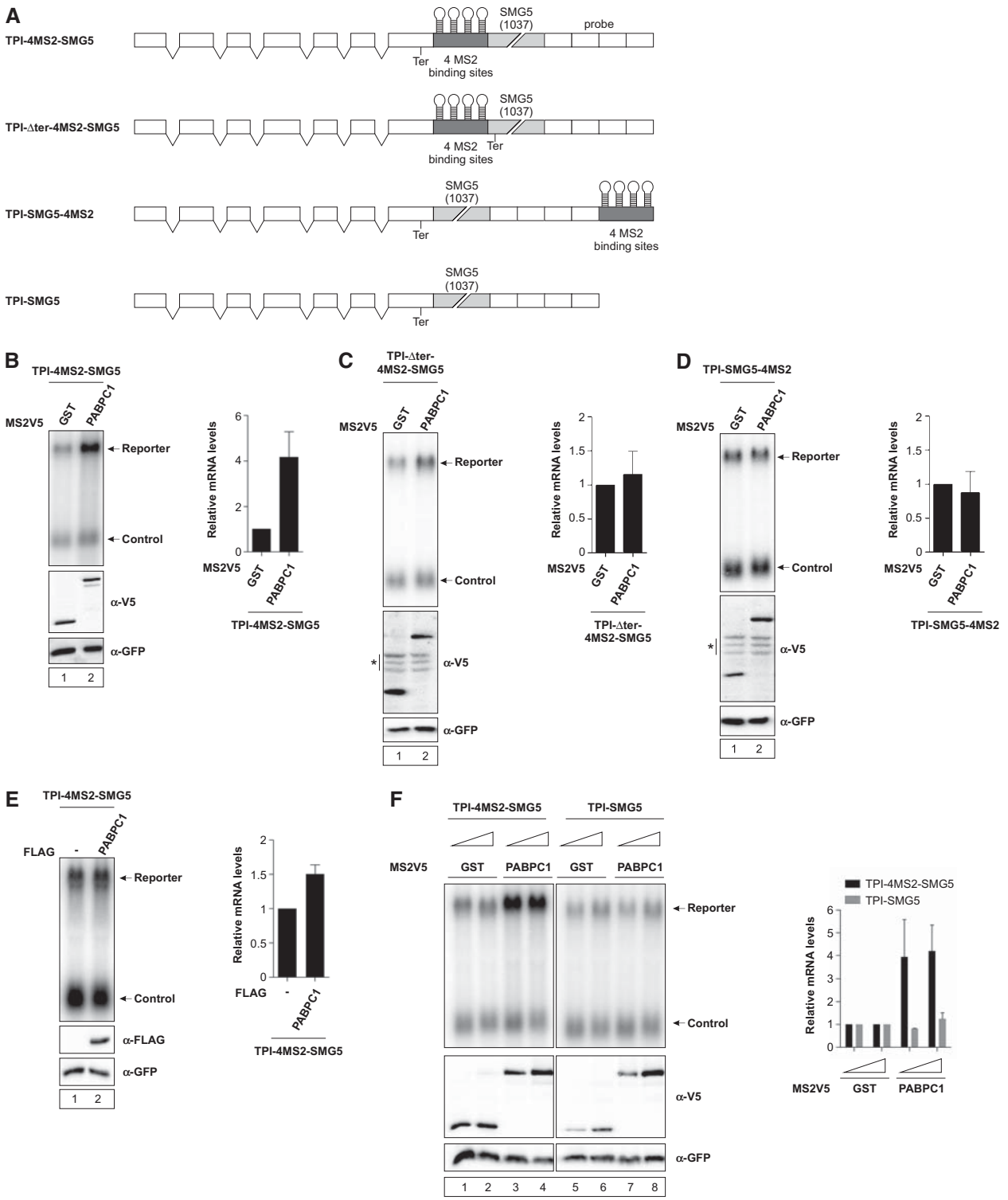


FIGURE 1. Tethering PABPC1 to a reporter mRNA containing a long 3′ UTR increases mRNA abundance. (*A*) Schematic representation of the triosephosphate isomerase (TPI) reporter constructs. White boxes depict exons, introns are shown as two connecting black lines, and Northern probe binding sites as white boxes without intron lines. Gray boxes represent MS2-binding area with MS2-stem–loops shown in black. The SMG5 3′ UTR is depicted as a light-gray box and the length in nucleotides is shown in brackets. (*B–F*) Northern blot analysis of total RNA extracted from HeLa cells transfected with plasmids expressing the indicated TPI reporter mRNA and MS2V5- or FLAG-tagged fusion proteins. A β-globin construct was cotransfected as control. Protein expression was detected by immunoblotting with α-V5 or α-FLAG antibody. Cotransfected GFP served as a loading control. Asterisks indicate unspecific bands (*C*,*D*). mRNA levels were normalized to MS2V5-GST (*B–D*,*F*) or pCI-FLAG (*E*). Bars represent the mean values of mRNA levels±SD upon tethering MS2V5-GST or MS2V5-PABPC1 (*B–D*,*F*), or pCI-FLAG and FLAG-PABPC1 (*E*). Concentrations of MS2V5-tagged protein expressing plasmids were increased from 1 µg (*F*, lanes *1*,*3*,*5*,*7*) to 3 µg (*F*, lanes *2*,*4*,*6*,*8*).

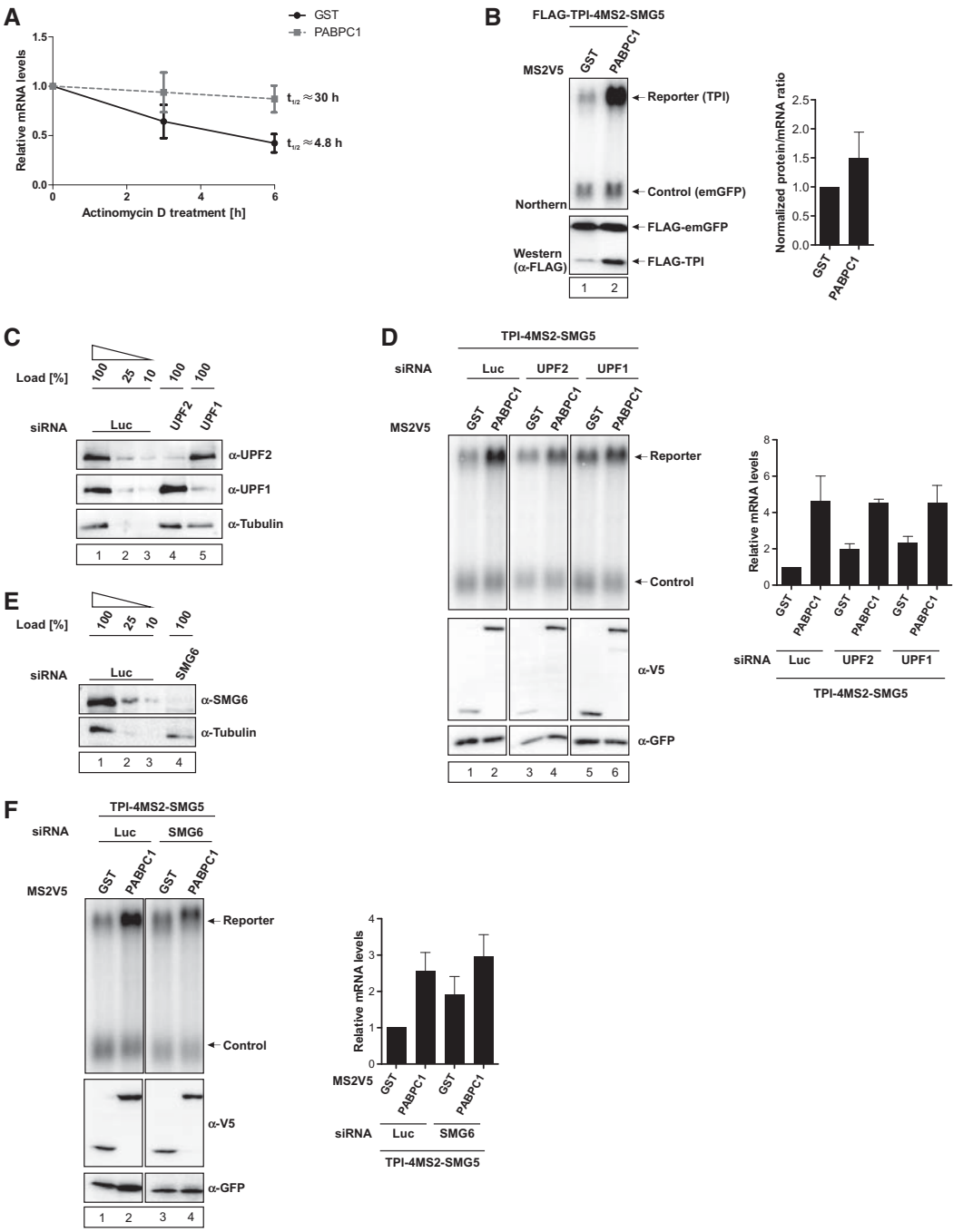


FIGURE 2. PABPC1 stabilizes reporter mRNA by suppressing NMD. (*A*) HeLa cells expressing reporter (TPI-4MS2-SMG5) and control mRNA, as well as MS2V5-tagged GST or PABPC1, were treated with Actinomycin D (5 µg/mL final concentration) for the indicated time prior to harvesting. Reporter mRNA levels were quantified by Northern blotting, normalized to control mRNA and GST control tethering, and plotted against time of Actinomycin D treatment. (*B*) Tethering of MS2V5-GST and -PABPC1 in HeLa cells cotransfected with N-terminally FLAG-tagged TPI-4MS2-SMG5 reporter and FLAG-tagged emGFP control expressing vectors. Both reporter and control mRNA contained heterologous binding sites in the 3′ UTR that enable the detection with the same Northern probe. Northern blot (*top*) and α-FLAG Western blot (*bottom*) analyses are shown. The signals for emGFP and TPI in Northern and Western blot experiments were quantified, normalized to the GST control lane, and the final ratio was calculated by normalizing protein expression levels to the respective mRNA expression levels. (*C*,*E*) siRNA-mediated knockdown of UPF2, UPF1, and SMG6. HeLa cells were transfected with siRNAs targeting UPF2, UPF1, SMG6, or Luciferase (negative control). The knockdown efficiency was assessed by immunoblotting with UPF2, UPF1, and SMG6-specific antibodies. Tubulin served as a loading control. (*D*,*F*) Northern blot analysis of TPI-4MS2-SMG5 reporter mRNA with β-globin mRNA as control in UPF2 (*D*, lanes *3*,*4*), UPF1 (*D*, lanes *5*,*6*), and SMG6 (*F*, lanes *3*,*4*) knockdown and control (*D*,*F*, lanes *1*,*2*) cells. HeLa cells were transfected with plasmids expressing MS2V5-GST and -PABPC1 proteins. Protein expression was detected by immunoblotting with a V5 antibody. GFP served as a loading control. mRNA levels were normalized to MS2V5-GST. Bars represent the mean values of mRNA levels ±SD upon tethering MS2V5-GST and -PABPC1 fusion proteins.

translation efficiency but rather by protecting the mRNA from degradation.

Human NMD is executed by a core machinery including the central NMD factors UPF1 and UPF2. To confirm that PABPC1 inhibits the canonical NMD pathway, we used small-interfering RNAs (siRNAs) to deplete UPF1 and UPF2 in human cell culture. Numerous studies have shown that NMD is impaired in cells lacking either of these two factors (Mendell et al. 2002; Gatfield et al. 2003). UPF1 and UPF2 were reduced to ~10% of regular expression levels by RNAi, as shown by immunoblotting (Fig. 2C). The levels of the reporter mRNAs were increased when UPF1 or UPF2 were depleted, confirming the inhibition of NMD by the transfected siRNAs (Fig. 2D, cf. lanes 1,3,5). Tethering of PABPC1 to the TPI-4MS2-SMG5 reporter mRNA increased reporter mRNA levels by a factor of more than four in control cells (Fig. 2D, lane 2), in line with our previous results. In contrast, tethering PABPC1 to the reporter only increased the mRNA levels by a factor of about two in both UPF2- and UPF1-knockdown cells (Fig. 2D, lanes 4,6). Notably, even though the change of mRNA levels by PABPC1 tethering was reduced in the UPF1 and UPF2 knockdown cells, the total levels of stabilized reporter mRNA under all three conditions reached a similar level (Fig. 2D, cf. lanes 2,4,6). These results indicate that PABPC1 acts as a strong suppressor in the canonical NMD pathway that involves the central NMD factors UPF1 and UPF2.

As described above, PABPC1 suppresses UPF1- and UPF2-dependent NMD. Several decay pathways act downstream from UPF1 to ensure efficient degradation of substrate mRNAs (Nicholson and Muhlemann 2010). We speculated that the NMD inhibiting effect of PABPC1 also impinges on the degradation phase of NMD. To test this hypothesis, we established the depletion of the NMD-specific endonuclease SMG6 by RNAi. In knockdown cells, the SMG6 protein levels were reduced to ~10% of regular expression levels (Fig. 2E). Similar to our results in UPF1- and UPF2-depleted cells, tethered PABPC1 stabilized the mRNA levels by a factor of less than two in SMG6-depleted cells (Fig. 2F, cf. lanes 2,4), while stabilized reporter levels remained unchanged compared with control knockdown cells. Taken together with the results obtained so far, this shows that PABPC1 indeed inhibits NMD by preventing the degradation of the substrate mRNA.

The interaction of PABPC1 and eRF3a is dispensable for NMD suppression

Current models of NMD suggest that PABPC1 competes with the NMD factor UPF1 for eRF3a binding (Singh et al. 2008). To more specifically elucidate which domains and interaction regions of PABPC1 are responsible for the NMD suppression effect observed in Figure 1, we generated six mutants of PABPC1 (Fig. 3A). PABPC1 interacts via its C-terminal MLLE domain with two PAM2 motifs present in the N terminus of eRF3a (Kozlov and Gehring 2010). We designed two mutants to impair this interaction: one consisting of RNA recognition motifs (RRMs) one to four of PABPC1 (PABPC1 RRM^{1234}) (Fig. 3A), i.e., lacking the C-terminal domain; the second containing a mutation of the MLLE motif to GAAR (PABPC1 $MLLE^{Mut}$) (Fig. 3A). The tethering assay shows that both PABPC1 RRM^{1234} and PABPC1 $MLLE^{Mut}$ suppressed NMD to almost the same extent as PABPC1 (Fig. 3B, lanes 2,3,5). We further tested two additional PABPC1 mutants abrogating PAM2 binding, one containing the MLLE mutation with additional point mutations known to further abolish PAM2 motif binding (Kozlov et al. 2004) (PABPC1 $MLLE^{Mut2}$) (Fig. 3A) and a longer version of RRM^{1234} including additional C-terminal amino acids (PABPC1 1-496) (Fig. 3A). Tethering either of these mutants to the TPI-4MS2-SMG5 reporter construct increases the mRNA abundance of the reporter to the same degree as PABPC1 (Fig. 3C). Notably, it was postulated that the direct binding of PABPC1 to eRF3a outcompetes the eRF3a-UPF1 interaction. However, our results suggest that this interaction, as well as other interactions involving the MLLE motif of PABPC1, is not strictly necessary for tethered PABPC1 to suppress NMD.

PABPC1 unable to bind eIF4G fails to stabilize reporter mRNA

The RRM^{1234} region of PABPC1 contains RRM2, which binds to a short N-terminal motif within eIF4G (Safaee et al. 2012). This interaction is important for mRNA circularization and the efficient expression of polyadenylated mRNAs (Wells et al. 1998; Amrani et al. 2008). Guided by the molecular structure of the PABPC1-eIF4G-RNA ternary complex (Safaee et al. 2012), we introduced two point mutations into PABPC1, which abolish binding to eIF4G ($PABPC1^{M161A/D165K}$) (Fig. 3A; Kahvejian et al. 2005). Strikingly, tethering of $PABPC1^{M161A/D165K}$ to the reporter no longer suppressed NMD and mRNA levels remained unchanged (Fig. 3B, lane 4). A similar result was obtained with the PABPC1 $MLLE^{Mut\ M161A/D165K}$ mutant that can neither interact with eIF4G nor eRF3a (Fig. 3B, lane 6). These results indicate that the interaction between PABPC1 and eIF4G is critical for the inhibition of NMD by tethered PABPC1.

To show that the results we have generated so far are indeed due to interactions or lack thereof between PABPC1 with eIF4G and eRF3a, we have performed in vitro interaction studies. Pull-down assays of C-terminally FLAG-tagged eRF3a with both of the PABPC1 $MLLE^{Mut}$ and PABPC1 1–496 mutants show that they are no longer able to interact with eRF3a, whereas the wild-type as well as the $PABPC1^{M161A/D165K}$ mutant are still able to be pulled down by eRF3a (Fig. 3D). The same experiment was performed with a shortened version of FLAG-tagged eIF4G (eIF4G 84–294) containing the PABPC1 binding site. In this experimental setup eIF4G 84–294 was no longer able to pull-down the $PABPC1^{M161A/D165K}$ mutant (Fig. 3E). The wild-type as

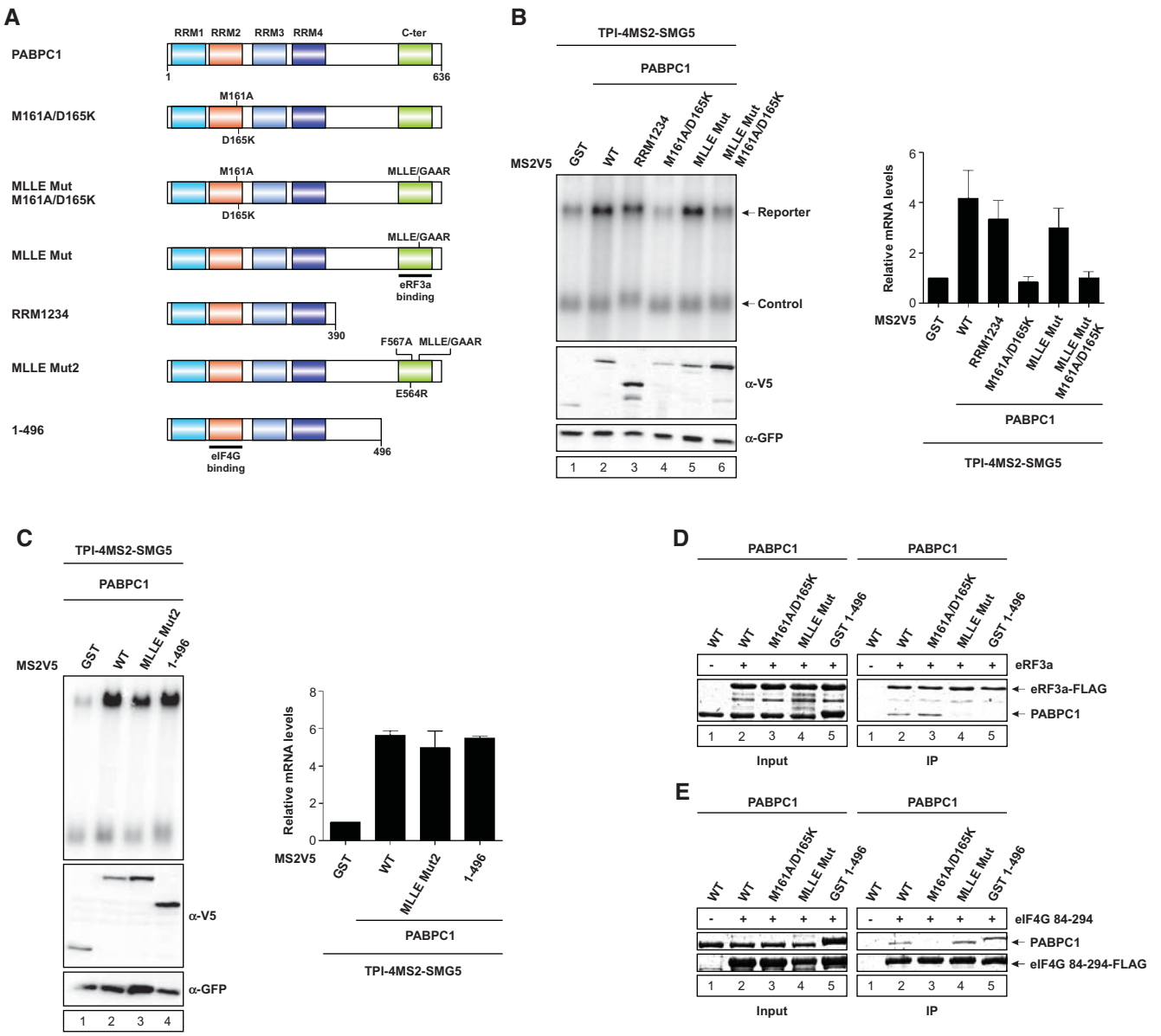


FIGURE 3. PABPC1 interaction with eIF4G, but not eRF3a, is essential for NMD suppression. (*A*) Schematic representation of PABPC1 domains and mutants. RNA recognition motifs (RRMs) and C-terminal region (C-ter) of PABPC1 are highlighted. Point mutations and binding sites are indicated. (*B*,*C*) HeLa cells were transfected with plasmids expressing the indicated MS2V5-tagged fusion proteins and the indicated TPI reporter mRNA. Northern blot analysis was performed and cotransfected β-globin mRNA construct served as control. Protein expression was detected by immunoblotting with a V5 antibody. GFP served as a loading control. mRNA levels were normalized to MS2V5-GST. Bars represent mean values of mRNA levels ±SD upon tethering of different MS2V5-tagged fusion proteins. (*D*,*E*) Pull-down assays of in vitro interaction studies using PABPC1 mutants and FLAG-tagged eRF3a or eIF4G 84-294. Proteins were visualized with Coomassie Brilliant Blue.

well as the PABPC1 MLLE^{Mut} and PABPC1 1–496 mutants were still able to interact with eIF4G 84–294 (Fig. 3E).

In summary, our results demonstrate that PABPC1 inhibits NMD when tethered upstream of an NMD-activating long 3′ UTR. Furthermore, binding of eIF4G but not eRF3a contributes to NMD suppression by tethered PABPC1. Interestingly, our findings suggest a previously unrecognized role of eIF4G-mediated mRNA circularization and ribosome recycling as modulators of NMD.

EJC-dependent NMD is mostly unaffected by PABPC1

So far, we have examined the suppression of NMD activated by the SMG5 3′ UTR. However, NMD of many nonsense-containing mRNAs occurs in a splicing-dependent manner and involves EJCs deposited at exon–exon junctions. Therefore, we wanted to analyze whether PABPC1 is able to antagonize EJC-dependent NMD. To this end, we

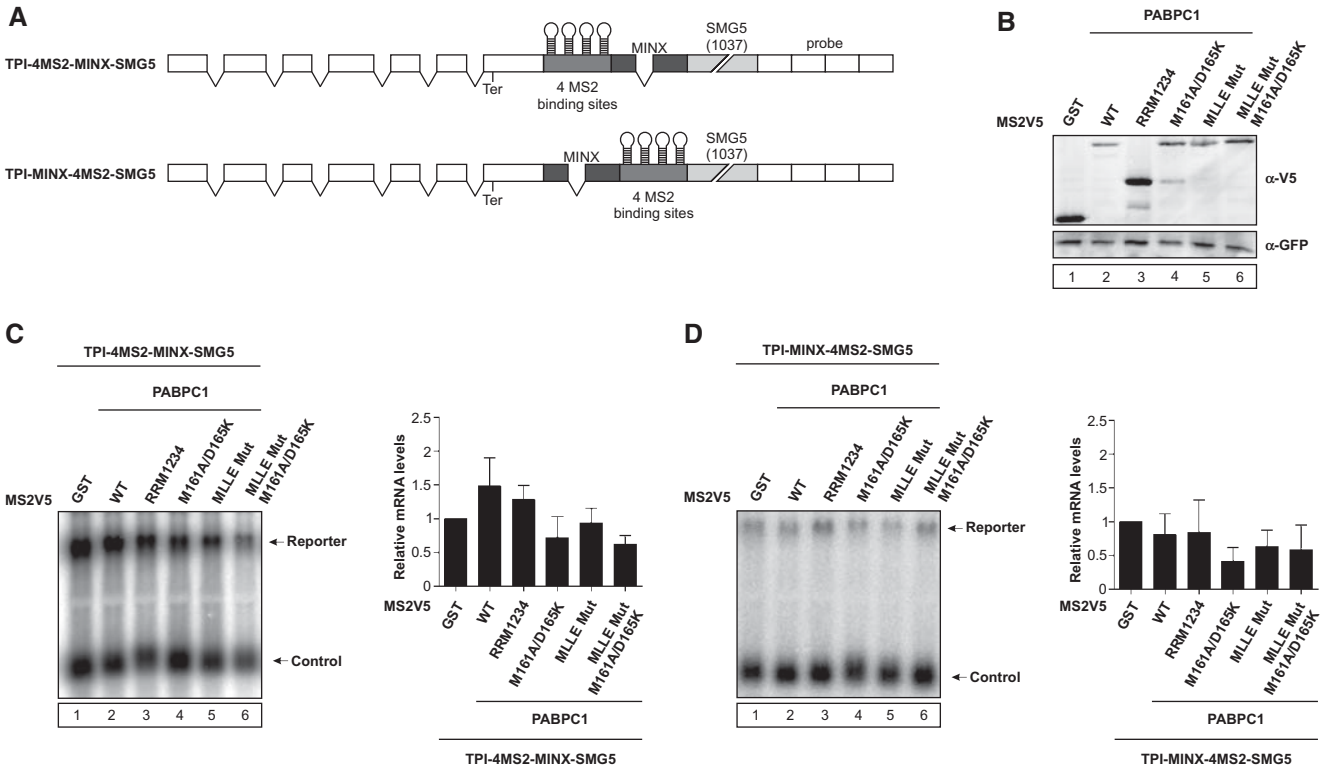


FIGURE 4. PABPC1 primarily functions as a suppressor in EJC-independent NMD. (*A*) Schematic representation of the TPI reporters as in Figure 1A. Dark-gray boxes indicate intron-containing MINX cassette. (*B*) Representative Western blot showing MS2V5-tagged protein expression, as analyzed by immunoblotting with a V5 antibody. (*C*,*D*) HeLa cells were transfected with plasmids expressing the indicated MS2V5-tagged fusion proteins and the indicated TPI reporter mRNA. Northern blot analysis was performed and cotransfected β-globin mRNA construct served as control. mRNA levels were normalized to MS2V5-GST. Bars represent mean values of mRNA levels ±SD upon tethering of different MS2V5-tagged fusion proteins.

constructed a reporter with an intron (MINX) downstream from the 4MS2-binding sites and upstream of the long 3′ UTR (TPI-4MS2-MINX-SMG5) (Fig. 4A). Splicing of the intron deposits an EJC that will activate NMD at the termination codon of the reporter mRNA. Strikingly, tethering of PABPC1 only weakly inhibited EJC-dependent NMD and slightly increased the MINX-containing mRNA levels by a factor of less than two (Fig. 4B,C, lane 2). In general, the inhibition of EJC-dependent NMD by PABPC1 variants was clearly reduced when compared with the inhibition of EJC-independent NMD (cf. Figs. 3B and 4C). Hence, the presence of an EJC appears to reduce the ability of PABPC1 to antagonize NMD, which explains why EJC-dependent NMD efficiently degrades mRNAs with short 3′ UTRs.

The weak inhibition of EJC-dependent NMD observed in Figure 4C was completely lost when we used a reporter construct, in which the MINX intron was inserted between the tethering sites and the termination codon (TPI-MINX-4MS2-SMG5) (Fig. 4A). Neither PABPC1 nor any of its mutants were able to increase mRNA levels of this reporter (Fig. 4B,D), which demonstrates that PABPC1 cannot antagonize the activation of NMD in the presence of an upstream EJC. Since an EJC in close proximity downstream from tethered PABPC1 is able to decrease the NMD suppression by PABPC1, we conclude that PABPC1 mainly regulates the EJC-independent NMD of mRNAs with long 3′ UTRs.

Tethered eRF3a relies on interaction with PABPC1 to antagonize NMD

To gain further insight into the role of eRF3a in NMD suppression, we tethered eRF3a to the TPI-4MS2-SMG5 reporter construct, which led to increased reporter mRNA abundance by a factor of four, demonstrating eRF3a's ability to suppress NMD similar to PABPC1 (Fig. 5A, lane 2). To investigate the importance of the interaction between eRF3a and PABPC1 in NMD suppression, we used an eRF3a mutant carrying a point mutation (eRF3a F76A), which is essential for binding to PABPC1 (Kononenko et al. 2010; Kozlov and Gehring 2010; Osawa et al. 2012). Compared with eRF3a WT, the eRF3a F76A mutant has lost its NMD suppressing activity (Fig. 5A, lane 3).

To ensure that the observed effects are in fact due to NMD suppression, we determined mRNA half-life when tethering eRF3a. Similar to PABPC1, eRF3a clearly increased mRNA half-life, confirming that it specifically suppresses NMD and stabilizes the reporter mRNA (Fig. 5B). Furthermore,

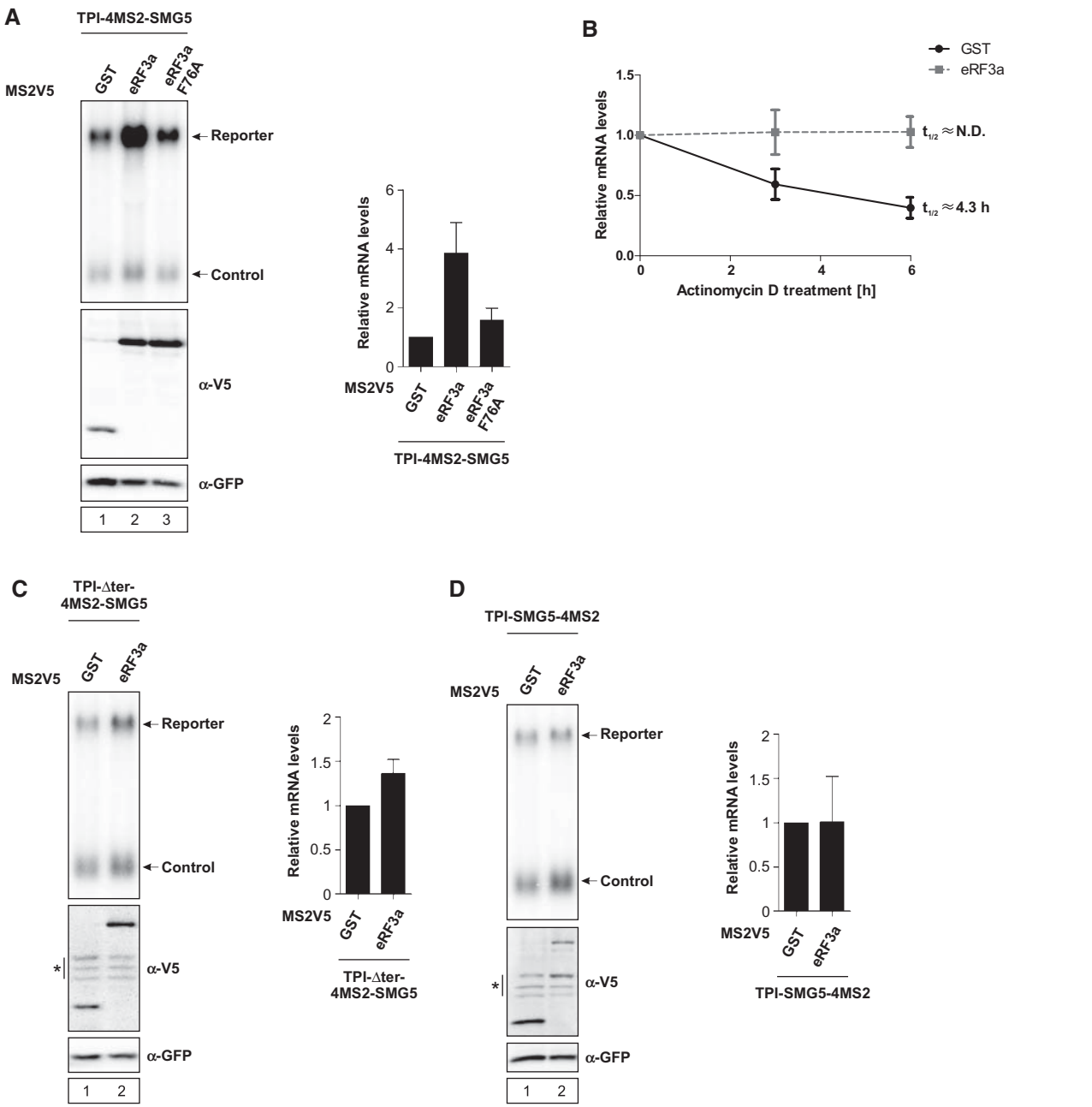


FIGURE 5. Tethered eRF3a unable to interact with PABPC1 fails to stabilize reporter mRNA levels. (*A*,*C*,*D*) HeLa cells were transfected with plasmids expressing the indicated MS2V5-tagged fusion proteins and the indicated TPI reporter mRNA. Northern blot analysis was performed and cotransfected β-globin mRNA construct served as control. Protein expression was detected by immunoblotting with a V5 antibody. GFP served as a loading control. Asterisks indicate unspecific bands (*C*,*D*). mRNA levels were normalized to MS2V5-GST. Bars represent mean values of mRNA levels ±SD upon tethering of different MS2V5-tagged fusion proteins. (*B*) HeLa cells expressing MS2V5-tagged GST or eRF3a were treated with Actinomycin D for the indicated time. mRNA levels were quantified after Northern blot analysis and plotted as described in Figure 2.

the specificity of the effects observed for tethered eRF3a were confirmed with both control reporter constructs described in Figure 1. Tethering eRF3a to either reporter did only marginally change mRNA abundance (Fig. 5C,D).

In summary, these results demonstrate that the interaction with PABPC1 is essential for eRF3a to suppress NMD initiated by a long 3′ UTR.

Recruitment of eIF4G increases mRNA abundance independently of interaction with PABPC1

Next, we aimed to further investigate NMD suppression by eIF4G and the role of the interaction between PABPC1 and eIF4G. An N-terminally truncated version of eIF4G (eIF4G

ΔN83) was used to analyze the function of eIF4G in NMD suppression (Fig. 6A).

Similar to PABPC1, we find that eIF4G ΔN83 is able to antagonize NMD, increasing the levels of the reporter mRNA by a factor of three (Fig. 6B, lane 2). To further elucidate the role of the interaction between eIF4G and PABPC1 in NMD suppression, we tethered different deletion mutants of eIF4G to the reporter construct (Fig. 6A). A shortened version of eIF4G ranging from amino acids 84 to 1089 (eIF4G 84–1089) (Fig. 6A) was still able to fully suppress NMD, which is indicated by an increase in mRNA reporter levels of a factor of three (Fig. 6B, lane 4). Surprisingly, a shortened version of eIF4G carrying a mutation of the KRERK motif at position 187–191 (eIF4G 84–1089 KRERK) (Fig. 6A), which is responsible for PABPC1 binding (Wakiyama et al. 2000), efficiently antagonized NMD (Fig. 6B, lane 5). Although the eIF4G KRERK mutant has previously been shown to abrogate binding between eIF4G and PABPC1 (Wakiyama et al. 2000), our pull-down assay indicates that the eIF4G KRERK mutant is still able to interact with PABPC1, albeit to a lesser degree (Fig. 6C). Hence, the full NMD suppression activity of tethered eIF4G 84–1089 KRERK might be due to partially retained PABPC1-binding. However, an eIF4G deletion mutant lacking the PABPC1-binding domain (eIF4G 206–1089) (Fig. 6A) had a slightly decreased NMD suppression rate, but was still able to increase mRNA abundance by a factor of two (Fig. 6B, lane 3). These results indicate that the interaction of eIF4G with PABPC1 is favorable but not absolutely necessary for eIF4G's ability to suppress NMD.

mRNA half-life was measured to ensure that the observed effects are in fact due to NMD suppression. Similar to PABPC1 and eRF3a, but less efficiently, eIF4G ΔN83 increased mRNA half-life, confirming that it specifically suppresses NMD and stabilizes the reporter mRNA (Fig. 6D). Notably, the low expression levels of the eIF4G constructs may account for the less prominent stabilization effects compared with tethered PABPC1 and eRF3a.

We confirmed the specificity of the effects observed for tethered eIF4G with both control reporter constructs described in Figure 1. Tethering eIF4G to either reporter only marginally changed mRNA abundance (Fig. 6E–G).

The results presented here indicate that suppression of NMD by tethered eIF4G does not strictly require binding to PABPC1, albeit this interaction might enhance the function of eIF4G.

DISCUSSION

While the process and the function of NMD has been elucidated in great detail and many factors and determinants that activate NMD are known, the mechanism of NMD suppression is far less understood. In this study, we report that the inhibition of EJC-independent NMD requires the interaction of PABPC1 with the initiation factor eIF4G. We further show that PABPC1 suppresses a SMG6-dependent canonical NMD pathway involving the central NMD factors UPF1 and UPF2. Our data suggest that the molecular processes of translation termination and mRNA circularization impinge on the activation of NMD.

The NMD-inhibitory function of PABPs has been observed in different eukaryotic organisms, such as yeast (Amrani et al. 2004), fly (Behm-Ansmant et al. 2007), and humans (Eberle et al. 2008; Ivanov et al. 2008; Silva et al. 2008; Singh et al. 2008). However, the NMD-specific function of PABP has proven to be difficult to study separately from its other important functions in mRNA translation and stabilization. Furthermore, human NMD is characterized by different signals for NMD activation and multiple pathways for degradation, contrary to many other organisms that use one main NMD mechanism (Hwang and Maquat 2011).

Human NMD occurs either in an EJC-dependent manner, when a PTC is located upstream of the last intron position, or independently of EJCs at termination codons that are followed by long 3′ UTRs (Schweingruber et al. 2013). We used a reporter mRNA with a long 3′ UTR to study the suppression of NMD by PABPC1. Our results demonstrate that tethering PABPC1 in close proximity of the termination codon impaired NMD of the reporter mRNA. We also confirmed that this effect is NMD-specific and does not occur on mRNAs with the tethering sites downstream from the 3′ UTR, without direct binding of PABPC1, or when the termination codon is moved to a position downstream from the tethering sites. Notably, the siRNA-mediated depletion of UPF1 or UPF2 is epistatic to NMD suppression by PABPC1, demonstrating that PABPC1 acts in the same pathway as, but antagonistically to, UPF1 and UPF2. We also provide evidence that PABPC1 antagonizes the SMG6-dependent degradation pathway that initiates NMD by endocleavage in the vicinity of the termination codon. Since we have not investigated the role of other NMD factors, it will remain an important challenge for future studies to test whether the suppression of NMD affects additional degradation pathways, such as SMG5/SMG7-dependent deadenylation or mRNA decapping.

Although previous studies suggested that PABPC1 also antagonizes EJC-dependent NMD (Ivanov et al. 2008; Singh et al. 2008), in our experiments the NMD inhibition by PABPC1 is reduced in the presence of a downstream EJC. This suggests that the NMD of the reporter mRNA without 3′ UTR introns occurs in an EJC-independent manner and does not involve EJCs bound at noncanonical positions within the long 3′ UTR (Sauliere et al. 2012; Singh et al. 2012). We interpret the weak inhibition of EJC-dependent NMD by PABPC1 as a nonspecific effect that is comparable to the effects observed with unrelated reporter mRNAs. However, it is conceivable that the specific composition of EJCs determines their amenability to NMD suppression or that a subset of EJC components are inhibited by PABPC1. Hence, it remains to be determined whether PABPC1 is able to antagonize

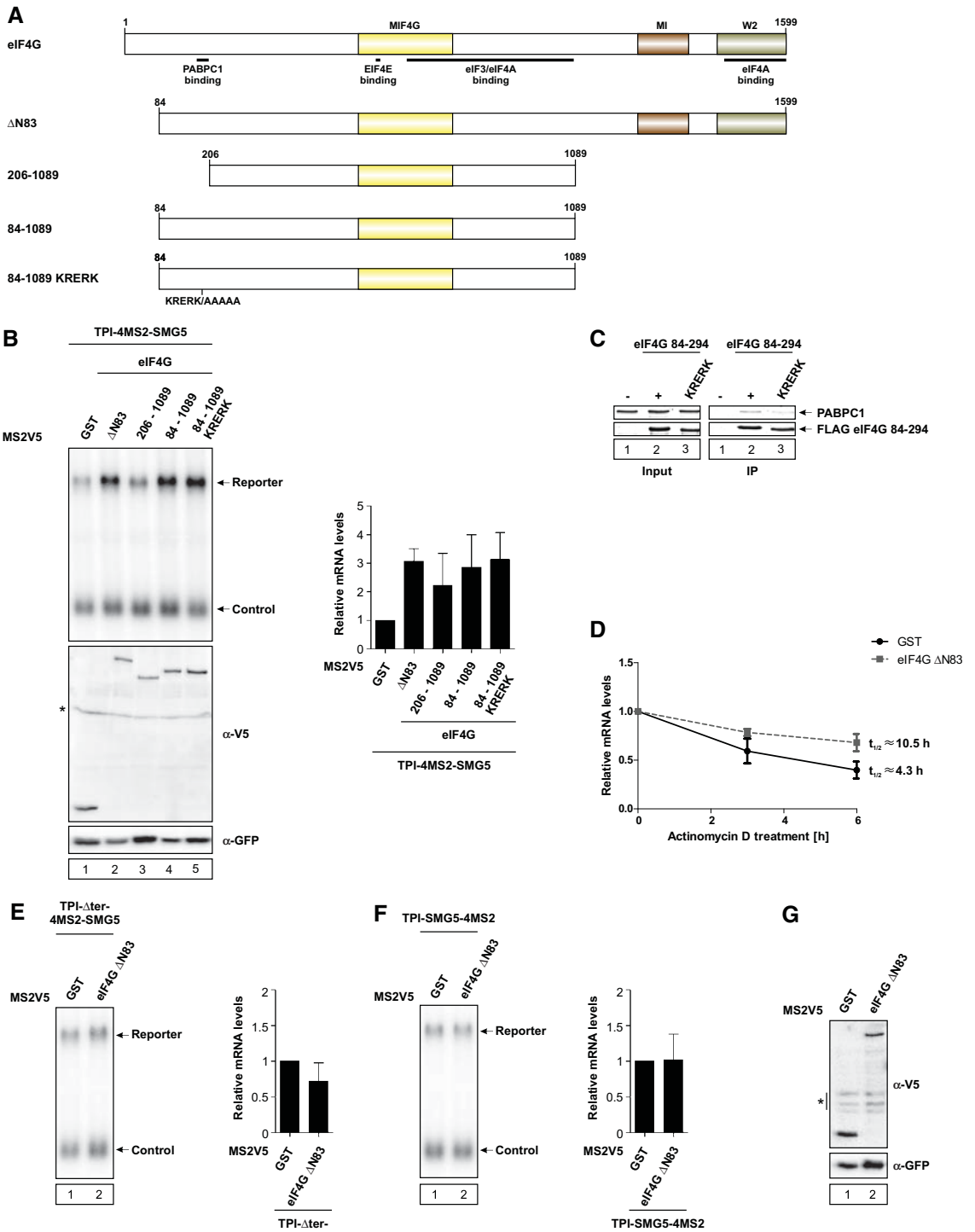


FIGURE 6. The interaction with PABPC1 is dispensable for NMD suppression by tethered eIF4G. (*A*) Illustration of eIF4G domain architecture and mutants. MIF4G, MI, and W2 domains are highlighted. Point mutations and binding sites are indicated. (*B*,*E*,*F*) HeLa cells were transfected with plasmids expressing the indicated MS2V5-tagged fusion proteins and the indicated TPI reporter mRNA. Northern blot analysis was performed and cotransfected β-globin mRNA construct served as control. Protein expression was detected by immunoblotting with a V5 antibody. GFP served as a loading control. Asterisk indicates an unspecific band (*B*). mRNA levels were normalized to MS2V5-GST. Bars represent mean values of mRNA levels ±SD upon tethering of different MS2V5-tagged fusion proteins. (*C*) Pull-down assays of in vitro interaction studies using PABPC1 and FLAG-tagged eIF4G 84-294 WT or KRERK mutant. Proteins were visualized with Coomassie Brilliant Blue. (*D*) HeLa cells expressing MS2V5-tagged GST or eIF4G ΔN83 were treated with Actinomycin D for the indicated time. mRNA levels were quantified after Northern blot analysis and plotted as described in Figure 2. (*G*) Representative Western blot showing MS2V5-tagged protein expression, as analyzed by immunoblotting with a V5 antibody. GFP served as a loading control. Asterisk indicates an unspecific band.

NMD activated by EJCs or other mRNA-bound proteins, or whether its function is restricted to long 3′ UTRs.

Current models of NMD suggest that a translation termination event in the proximity of the poly(A) tail prevents the interaction of UPF1 with eRF3a and therefore inhibits NMD, whereas translation termination upstream of a long 3′ UTR enables the association of UPF1 with eRF3a and activates NMD. In contrast to this model, we find that tethered PABPC1 does not require the interaction with eRF3a, but needs to bind to eIF4G to suppress NMD. Similar to PABPC1, tethered eIF4G also inhibits NMD. Additionally, mutants of eIF4G lacking the binding region for PABPC1 either by deletion or point mutation are still able to suppress NMD. These results suggest that binding of PABPC1 to eIF4G is at least partially dispensable for NMD suppression by eIF4G. We suggest that tethering of eIF4G downstream from the termination codon establishes a link of the site of translation termination to the 5′ end of the mRNA and facilitates ribosome recycling in a PABPC1-independent manner. Surprisingly, we also observe a strong suppression of NMD by tethered eRF3a. This result was unexpected in light of our PABPC1 results, which demonstrate that eRF3a is dispensable for NMD suppression by PABPC1. However, we hypothesize that tethering of eRF3a enhances ribosome recycling at the termination codon by the recruitment of PABPC1, which explains the inhibition of NMD by eRF3a. This hypothesis is supported by our observation that a mutant of eRF3a (F76A), which is unable to interact with PABPC1, is no longer able to inhibit NMD in the tethering assay. Of note, additional binding partners of PABPC1 known to compete with the interaction of eIF4G and eRF3a may be involved in the regulation of NMD suppression, but were not studied here. Furthermore, PABPC4 and eRF3b may confer tissue-specific NMD suppression comparable to their homologs PABPC1 and eRF3a, respectively (Chauvin et al. 2005; Burgess et al. 2011). In summary, we have started to map NMD-suppressing domains of eIF4G and eRF3a, but a precise identification of critical interactions will be required to delineate the network of proteins that contribute to NMD suppression.

The interaction of PABPC1 with eIF4G is thought to facilitate circularization of mRNAs, support ribosome recycling, and initiate translation (Wells et al. 1998; Kahvejian et al. 2005; Amrani et al. 2008). These processes, albeit being important for general translation, have not been previously linked to NMD. Hence, we suggest a revised model of NMD to include our findings (Fig. 7). According to this model, closed loop formation of the mRNA via PABPC1-eIF4G is important not only for translation initiation and ribosome recycling, but also for the suppression of NMD (Fig. 7A,C). We suggest that the interaction of PABPC1 with eRF3a establishes a branch connection from the site of translation termination to the 5′ end of the mRNA through the eIF4G-PABPC1 binding (Jackson et al. 2010). While the interaction of PABPC1 with eIF4G stimulates the removal

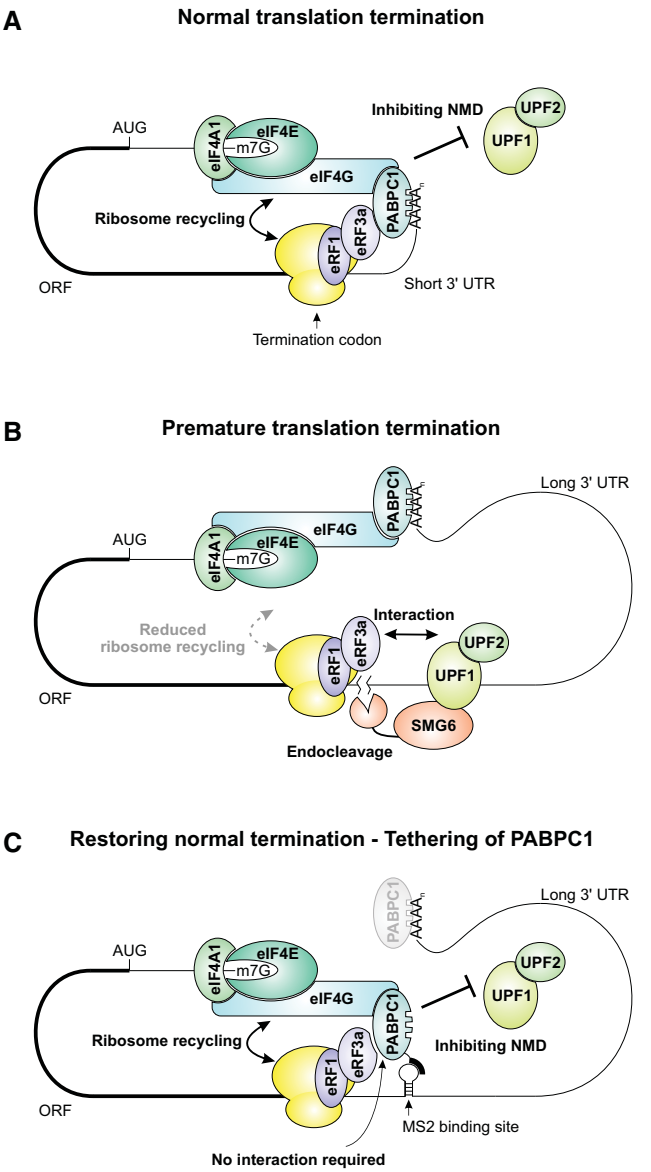


FIGURE 7. Model for a link between eIF4G-mediated ribosome recycling and NMD inhibition. (*A*) Normal translation termination and ribosome recycling of a short 3′ UTR-containing mRNA is enabled by PABPC1 interacting with both eRF3a and eIF4G, thereby preventing NMD. (*B*) Aberrant translation termination of long 3′ UTR-containing mRNAs activates NMD. The interaction of PABPC1 with eRF3a is decreased due to a large physical distance, preventing efficient ribosome recycling by eIF4G. Consequently, UPF1 is postulated to interact with eRF3a and elicit NMD. (*C*) Tethered PABPC1 inhibits NMD of a long 3′ UTR-containing substrate by bringing eIF4G in close proximity to the terminating ribosome. This proximity promotes a proper translation termination event and facilitates ribosome recycling, thereby antagonizing NMD activation.

of ribosomes from the site of termination, the inability to remove a ribosome after termination serves as an NMD activating signal and enables the recruitment of the NMD machinery to the position of the termination codon (Fig. 7B). This is in line with previous reports implicating that PTCs

differ from normal termination codons most likely in the rate of ribosome dissociation subsequent to peptide hydrolysis (Kervestin and Jacobson 2012). We propose that the inability to recycle ribosomes is a stochastic event that can occur during every round of translation termination and no difference whatsoever is expected between different modes of translation initiation (Durand and Lykke-Andersen 2013; Rufener and Muhlemann 2013). While it is conceivable that many aberrant mRNAs are efficiently recognized during the first termination event, the correct position of the termination codon will be monitored during every translation cycle and serves to eliminate mRNAs that once escaped decay. Hence, the continuous recycling of ribosomes via mRNA circularization at the termination codon may represent a main mechanism to prevent the degradation of mRNAs with short 3′ UTRs. The precise molecular signal that recruits the surveillance complex to the site of termination still remains to be determined. UPF1 likely represents the factor that directly interprets the signals of terminating ribosomes, either via a direct interaction with the ribosome (Min et al. 2013) or by communicating with the eukaryotic release factors (Ivanov et al. 2008; Singh et al. 2008). However, it is unclear how phosphorylation of UPF1 is activated in the absence of a downstream EJC (Kashima et al. 2006). While our work complements the current model of NMD with an important role of the eRF3a-PABPC1-eIF4G interaction, the molecular function of many additional factors involved in NMD, such as components of the EJC, still needs to be integrated.

MATERIALS AND METHODS

Plasmids constructs

Plasmid constructs β-globin, WT300+e3, pCI-FLAG, pCI-MS2V5, pCI-mVenus, pCI-TPI, and expression vectors for PABPC1 and eIF4G were described previously (Gehring et al. 2005, 2009; Hundsdoerfer et al. 2005; Ivanov et al. 2008; Steckelberg et al. 2012). The modification of pCI-TPI with four copies of binding sites for the heterologous probe used in Northern blot analysis, as well as the insertion of the SMG5 3′ UTR and 4MS2 binding sites in the 3′ UTR of TPI were described elsewhere (V Boehm, N Haberman, F Ottens, J Ule, and NH Gehring, in prep.). Using the same cassette cloning strategy, the MINX splicing cassette was introduced in the vectors (TPI-4MS2-MINX-SMG5, TPI-MINX-4MS2-SMG5). Deletion and point mutants of PABPC1, eRF3a and eIF4G were generated by PCR, cloned in the designated expression vector, and verified by sequencing.

siRNA transfections

HeLa cells were grown in 6-cm plates and transiently transfected with 300 pmol siRNA using Lipofectamine RNAiMAX (Life Technologies). At 24 h post-transfection the cells were transferred to 10-cm plates and 1 d later transfected again with 600 pmol siRNA. The following siRNA target sequences were used for Luciferase 5′-AACGTACGCGGAATACTTCGA-3′, for SMG6 5′-AAGGGTCACAGTGCTGAAGTA-3′, for UPF1 5′-AAGATGCAGTTCCGCTCCATT-3′, and for UPF2 5′-CACGTTGTGGATGGAGTGTTA-3′.

Plasmid transfections

HeLa cells were grown in 6-well plates or transferred 1 d after siRNA transfection to 6-well plates and transfected by calcium phosphate precipitation with 0.3 µg of a mVenus expression plasmid, 0.5 µg control plasmid (β-globin or WT300+e3), and 2 µg plasmid encoding for reporter mRNA. For tethering or overexpression of tagged-protein, 0.8 µg of the FLAG- or MS2V5 expression plasmid was included in the transfection mix. For mRNA half-life experiments, cells were incubated with medium supplemented with 5 µg/mL Actinomycin D for 3 or 6 h prior to harvesting.

RNA extraction and Northern blotting

Total RNA was extracted with Isol-RNA Lysis Reagent (5PRIME) and analyzed by Northern blotting as described (Gehring et al. 2009). Signals were quantified using a Typhoon Trio (GE Healthcare).

Immunoblot analysis

SDS-PAGE and immunoblot analysis was performed using protein samples derived from Isol-RNA Lysis Reagent extractions. The antibodies against tubulin (T6074) and FLAG (F7425) were from Sigma, the antibody against V5 (18870) was from QED Bioscience, the antibodies against GFP (ab290) and SMG6 (ab87539) were from Abcam, and the antibodies against UPF1 and UPF2 were kindly provided by Jens Lykke-Andersen.

Expression and purification of recombinant proteins

Strep-tagged PABPC1 wild type and mutants, GST-tagged PABPC1 1–496 and N-terminally GST-, C-terminally FLAG-tagged eRF3a and eIF4G constructs were expressed in *E. coli* Rosetta 2. Cells were grown to exponential phase in LB medium (OD_{600} = 0.6–0.8) and expression was induced with 0.2 mM IPTG overnight at 20° C. Strep-tagged proteins were purified via affinity chromatography using StrepTactin Superflow Plus columns (Qiagen). GST-tagged proteins were purified via affinity chromatography using GSTrap columns (GE Healthcare) followed by size exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare). Cell lysis was performed in 40 mM Tris (pH 7.8), 250 mM NaCl with protease inhibitors (Protease inhibitor cocktail [Sigma], and 1 mM PMSF). All constructs were stored in 40 mM Tris (pH 7.8) and 150 mM NaCl.

In vitro pull-down analysis

Three hundred picomoles of FLAG-tagged proteins (eIF4G 84-294 or eRF3a) were incubated with 250 pmol PABPC1 constructs in a final volume of 400 µL binding buffer (25 mM HEPES at pH 7.8; 150 mM NaCl; 2 mM $MgCl_2$; 0.1% NP-40; 0.01% Triton X-100) in the presence of magnetic beads coupled to anti-FLAG antibodies (M2 magnetic beads; Sigma). After incubation for 2 h at 4°C, beads were washed twice with 500 µL wash buffer (25 mM HEPES at pH

7.8; 300 mM NaCl; 2 mM $MgCl_2$; 0.1% NP-40; 0.2% Triton X-100) and coprecipitated proteins were eluted with 1x SDS loading buffer. 10% of the protein mix was used as input control, all samples were separated on 12% SDS–polyacrylamide gels and stained with Coomassie Brilliant Blue.

ACKNOWLEDGMENTS

We thank Heidi Thelen and Juliane Hancke for excellent technical assistance; the Leptin, Schnetz, and Uhlirova labs for sharing equipment; Gabriele Neu-Yilik and members of the Gehring lab for useful discussions. We are grateful to Jens Lykke-Andersen for antibodies against UPF1 and UPF2 and Matthias Hentze and Dunja Ferring-Appel for the eIF4G expression plasmid. V.B. is supported by a fellowship from the International Graduate School in Development Health and Disease. This research was funded by grants from the Fritz Thyssen Stiftung and the Deutsche Forschungsgemeinschaft (SFB635, B6) to N.H.G.

Received February 20, 2014; accepted June 6, 2014.

REFERENCES

Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A. 2004. A faux 3′-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432:** 112–118.

Amrani N, Ghosh S, Mangus DA, Jacobson A. 2008. Translation factors promote the formation of two states of the closed-loop mRNP. *Nature* **453:** 1276–1280.

Behm-Ansmant I, Gatfield D, Rehwinkel J, Hilgers V, Izaurralde E. 2007. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J* **26:** 1591–1601.

Bhuvanagiri M, Schlitter AM, Hentze MW, Kulozik AE. 2010. NMD: RNA biology meets human genetic medicine. *Biochem J* **430:** 365–377.

Burgess HM, Richardson WA, Anderson RC, Salaun C, Graham SV, Gray NK. 2011. Nuclear relocalisation of cytoplasmic poly(A)-binding proteins PABP1 and PABP4 in response to UV irradiation reveals mRNA-dependent export of metazoan PABPs. *J Cell Sci* **124:** 3344–3355.

Chang YF, Imam JS, Wilkinson MF. 2007. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* **76:** 51–74.

Chauvin C, Salhi S, Le Goff C, Viranaicken W, Diop D, Jean-Jean O. 2005. Involvement of human release factors eRF3a and eRF3b in translation termination and regulation of the termination complex formation. *Mol Cell Biol* **25:** 5801–5811.

Clerici M, Deniaud A, Boehm V, Gehring NH, Schaffitzel C, Cusack S. 2013. Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2. *Nucleic Acids Res* **42:** 2673–2686.

Durand S, Lykke-Andersen J. 2013. Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat Struct Mol Biol* **20:** 702–709.

Eberle AB, Stalder L, Mathys H, Orozco RZ, Muhlemann O. 2008. Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. *PLoS Biol* **6:** e92.

Eberle AB, Lykke-Andersen S, Muhlemann O, Jensen TH. 2009. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* **16:** 49–55.

Gatfield D, Unterholzner L, Ciccarelli FD, Bork P, Izaurralde E. 2003. Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. *EMBO J* **22:** 3960–3970.

Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH, Hentze MW, Kulozik AE. 2005. Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol Cell* **20:** 65–75.

Gehring NH, Lamprinaki S, Kulozik AE, Hentze MW. 2009. Disassembly of exon junction complexes by PYM. *Cell* **137:** 536–548.

Glavan F, Behm-Ansmant I, Izaurralde E, Conti E. 2006. Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J* **25:** 5117–5125.

Hoshino S, Imai M, Kobayashi T, Uchida N, Katada T. 1999. The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3′-poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. *J Biol Chem* **274:** 16677–16680.

Hundsdoerfer P, Thoma C, Hentze MW. 2005. Eukaryotic translation initiation factor 4GI and p97 promote cellular internal ribosome entry sequence-driven translation. *Proc Natl Acad Sci* **102:** 13421–13426.

Huntzinger E, Kashima I, Fauser M, Sauliere J, Izaurralde E. 2008. SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* **14:** 2609–2617.

Hwang J, Maquat LE. 2011. Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: To die or not to die, that is the question. *Curr Opin Genet Dev* **21:** 422–430.

Ivanov PV, Gehring NH, Kunz JB, Hentze MW, Kulozik AE. 2008. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J* **27:** 736–747.

Jackson RJ, Hellen CU, Pestova TV. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11:** 113–127.

Kahvejian A, Svitkin YV, Sukarieh R, M'Boutchou MN, Sonenberg N. 2005. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev* **19:** 104–113.

Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S. 2006. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev* **20:** 355–367.

Kervestin S, Jacobson A. 2012. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* **13:** 700–712.

Kononenko AV, Mitkevich VA, Atkinson GC, Tenson T, Dubovaya VI, Frolova LY, Makarov AA, Hauryliuk V. 2010. GTP-dependent structural rearrangement of the eRF1:eRF3 complex and eRF3 sequence motifs essential for PABP binding. *Nucleic Acids Res* **38:** 548–558.

Kozlov G, Gehring K. 2010. Molecular basis of eRF3 recognition by the MLLE domain of poly(A)-binding protein. *PLoS One* **5:** e10169.

Kozlov G, De Crescenzo G, Lim NS, Siddiqui N, Fantus D, Kahvejian A, Trempe JF, Elias D, Ekiel I, Sonenberg N, et al. 2004. Structural basis of ligand recognition by PABC, a highly specific peptide-binding domain found in poly(A)-binding protein and a HECT ubiquitin ligase. *EMBO J* **23:** 272–281.

Loh B, Jonas S, Izaurralde E. 2013. The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* **27:** 2125–2138.

Mendell JT, ap Rhys CM, Dietz HC. 2002. Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. *Science* **298:** 419–422.

Min EE, Roy B, Amrani N, He F, Jacobson A. 2013. Yeast Upf1 CH domain interacts with Rps26 of the 40S ribosomal subunit. *RNA* **19:** 1105–1115.

Nicholson P, Muhlemann O. 2010. Cutting the nonsense: the degradation of PTC-containing mRNAs. *Biochem Soc Trans* **38:** 1615–1620.

Nicholson P, Yepiskoposyan H, Metze S, Zamudio Orozco R, Kleinschmidt N, Muhlemann O. 2010. Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell Mol Life Sci* **67:** 677–700.

Okada-Katsuhata Y, Yamashita A, Kutsuzawa K, Izumi N, Hirahara F, Ohno S. 2012. N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res* **40:** 1251–1266.

Osawa M, Hosoda N, Nakanishi T, Uchida N, Kimura T, Imai S, Machiyama A, Katada T, Hoshino S, Shimada I. 2012. Biological role of the two overlapping poly(A)-binding protein interacting motifs 2 (PAM2) of eukaryotic releasing factor eRF3 in mRNA decay. *RNA* **18:** 1957–1967.

Rebbapragada I, Lykke-Andersen J. 2009. Execution of nonsense-mediated mRNA decay: What defines a substrate? *Curr Opin Cell Biol* **21:** 394–402.

Rufener SC, Muhlemann O. 2013. eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat Struct Mol Biol* **20:** 710–717.

Safaee N, Kozlov G, Noronha AM, Xie J, Wilds CJ, Gehring K. 2012. Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol Cell* **48:** 375–386.

Sauliere J, Murigneux V, Wang Z, Marquenet E, Barbosa I, Le Tonqueze O, Audic Y, Paillard L, Roest Crollius H, Le Hir H. 2012. CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat Struct Mol Biol* **19:** 1124–1131.

Schweingruber C, Rufener SC, Zund D, Yamashita A, Muhlemann O. 2013. Nonsense-mediated mRNA decay—mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim Biophys Acta* **1829:** 612–623.

Silva AL, Ribeiro P, Inacio A, Liebhaber SA, Romao L. 2008. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA* **14:** 563–576.

Singh G, Rebbapragada I, Lykke-Andersen J. 2008. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol* **6:** e111.

Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ. 2012. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* **151:** 750–764.

Steckelberg AL, Boehm V, Gromadzka AM, Gehring NH. 2012. CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep* **2:** 454–461.

Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N. 2012. Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol* **9:** 1370–1379.

Thermann R, Neu-Yilik G, Deters A, Frede U, Wehr K, Hagemeier C, Hentze MW, Kulozik AE. 1998. Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J* **17:** 3484–3494.

Uchida N, Hoshino S, Imataka H, Sonenberg N, Katada T. 2002. A novel role of the mammalian GSPT/eRF3 associating with poly(A)-binding protein in cap/poly(A)-dependent translation. *J Biol Chem* **277:** 50286–50292.

Wakiyama M, Imataka H, Sonenberg N. 2000. Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for *Xenopus* oocyte maturation. *Curr Biol* **10:** 1147–1150.

Wells SE, Hillner PE, Vale RD, Sachs AB. 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell* **2:** 135–140.

Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Muhlemann O. 2011. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **17:** 2108–2118.

Zhang J, Sun X, Qian Y, Maquat LE. 1998. Intron function in the nonsense-mediated decay of β-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *RNA* **4:** 801–815.

The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay

Tobias Fatscher, Volker Boehm, Benjamin Weiche, et al.

RNA 2014 20: 1579-1592 originally published online August 21, 2014
Access the most recent version at doi:10.1261/rna.044933.114

Cell Reports
Report

CWC22 Connects Pre-mRNA Splicing and Exon Junction Complex Assembly

Anna-Lena Steckelberg,[1] Volker Boehm,[1] Agnieszka M. Gromadzka,[1] and Niels H. Gehring[1,*]
[1]University of Cologne, Institute for Genetics, Zuelpicher Str. 47a, D-50674 Cologne, Germany
*Correspondence: ngehring@uni-koeln.de
http://dx.doi.org/10.1016/j.celrep.2012.08.017

SUMMARY

The exon junction complex (EJC) is a key regulator of posttranscriptional mRNA fate and binds to mRNA during splicing. Although the composition of EJCs is well understood, the mechanism mediating splicing-dependent EJC assembly and the factor(s) recruiting the EJC remain elusive. Here, we identify CWC22 as an essential splicing factor that is required for EJC assembly. In CWC22-depleted cells, pre-mRNA splicing is impaired but is rescued by a central fragment of CWC22. We show that the MIF4G domain of CWC22 initiates EJC assembly via a direct interaction with the EJC core protein eIF4A3, and we characterize mutations in eIF4A3 that abolish binding to CWC22. These eIF4A3 mutants efficiently nucleate splicing-independent recombinant EJC core complexes, but they fail to support splicing-dependent EJC deposition. Our work establishes a direct link between the splicing machinery and the EJC, hence uncovering a molecular interaction at the center of a posttranscriptional gene regulation network.

INTRODUCTION

The molecular apparatus required for eukaryotic gene expression forms a network that is able to communicate across the nuclear envelope and includes the transcription, splicing, and translation machinery (Moore and Proudfoot, 2009). Interactions between the different processing steps are mediated by RNA-binding proteins that form messenger ribonucleoprotein complexes (mRNPs) (Rodríguez-Navarro and Hurt, 2011). The protein composition of the mRNP reflects the history and determines the fate of each mRNA molecule. A key regulator within this mRNP network is the exon junction complex (EJC).

The EJC is a protein complex that assembles on spliced mRNA 20–24 nts upstream of exon-exon boundaries in a sequence-independent manner (Le Hir et al., 2000). It remains bound to the mRNA, thereby marking the position of splice sites, until it is disassembled during translation in the cytoplasm (Gehring et al., 2009b). In mammals, EJCs are involved in the detection of mRNAs containing premature translation termination codons (Lykke-Andersen et al., 2001). Furthermore, components of the EJC have been shown to be responsible for the increased export and translational efficiency of spliced mRNAs (Le Hir et al., 2001; Ma et al., 2008). They also regulate the splicing of a subset of pre-mRNAs, particularly those containing long introns, such as mitogen-activated protein kinase (MAPK) (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010).

The EJC core consists of four subunits, eIF4A3 (DDX48), MAGOH, Y14 (RBM8A), and BTZ (CASC3, MLN51), of which MAGOH and Y14 constitute a stable heterodimer (Tange et al., 2005). The crystal structure of the EJC core reveals how these subunits are organized into a complex (Andersen et al., 2006; Bono et al., 2006). In the presence of ATP, the DEAD-box helicase eIF4A3 adopts a closed conformation and binds to RNA. The closed conformation and therefore the RNA interaction is stabilized upon binding of the MAGOH-Y14 heterodimer to eIF4A3 (Ballut et al., 2005). The core component BTZ binds to eIF4A3 in a conformation-independent manner, although the interaction is stabilized within the EJC by additional protein-protein and protein-RNA interactions. Although the EJC can form spontaneously in vitro from recombinant components, its assembly in living cells or cell extracts is strictly splicing dependent (Bono and Gehring, 2011). However, the underlying molecular mechanism of EJC loading, the reason for the splicing dependence, and the spliceosomal proteins involved in EJC assembly remain unknown.

Here, we show that the protein CWC22 is required for pre-mRNA splicing in mammalian cells and represents an essential EJC loading factor. CWC22 interacts directly with eIF4A3 via its MIF4G domain, and this interaction contributes to the splicing-dependent assembly of EJCs. In cells depleted of CWC22, pre-mRNA splicing is impaired, while a construct comprising the MIF4G and MA3 domain of CWC22 suffices to rescue the splicing defect. Taken together, our results highlight the functional interconnection between splicing and EJC recruitment and indicate that eIF4G-eIF4A-like interactions emerge as common principle in mRNP remodeling.

RESULTS

CWC22 Is an eIF4A3-Interacting Protein

The EJC core protein eIF4A3 directly contacts RNA and therefore serves as an interaction platform for all other EJC proteins (Andersen et al., 2006; Bono et al., 2006). In order to identify putative factors within the spliceosome that recruit eIF4A3 and initiate EJC assembly, we searched the human genome

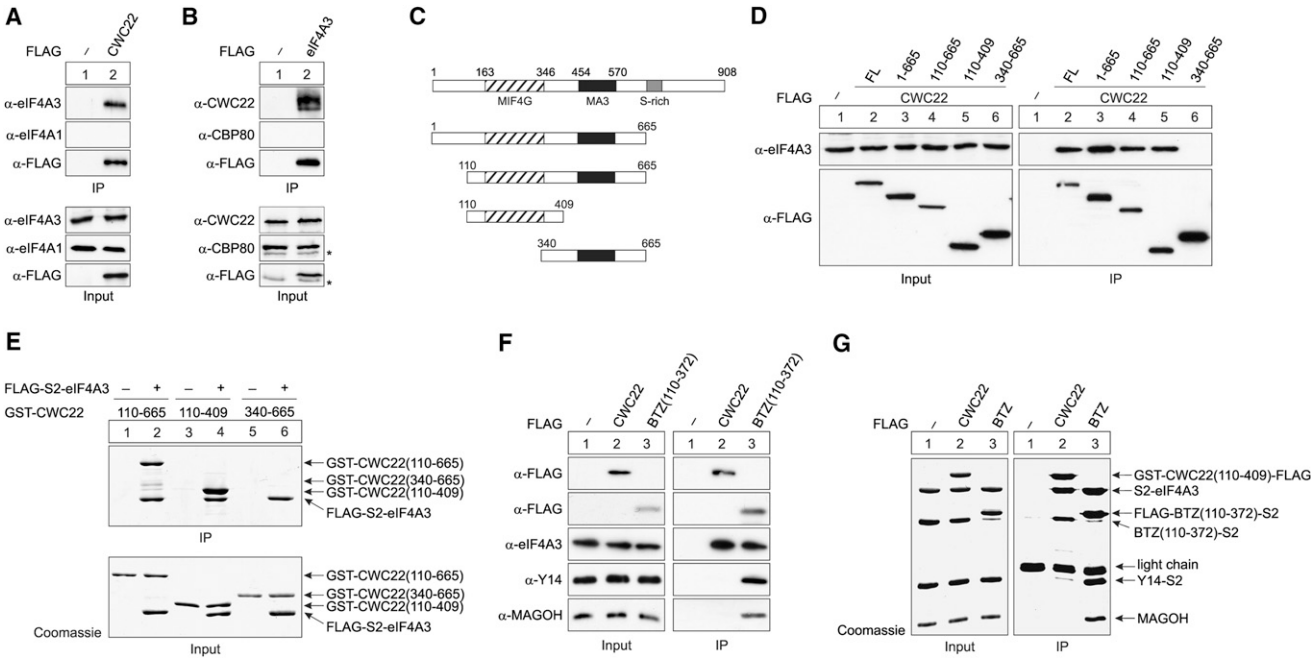


Figure 1. eIF4A3 Binds to CWC22 via Its MIF4G Domain

(A) FLAG-CWC22 was immunoprecipitated from RNase-A-treated cell lysates. Coprecipitated endogenous proteins were detected by immunostaining with specific antibodies. Unfused FLAG was used as a negative control.

(B) FLAG-eIF4A3 was immunoprecipitated as described in (A). Asterisks denote nonspecific bands.

(C) Schematic representation of N- and C-terminal truncations of CWC22 used for FLAG immunoprecipitation.

(D) FLAG-CWC22 deletion mutants were immunoprecipitated as described in (A).

(E) Immunoprecipitation of recombinant CWC22 and eIF4A3. FLAG-Strep-eIF4A3, GST-CWC22 (110–409),GST-CWC22 (110–665), and GST-CWC22 (340–665) were purified from *E.coli*, incubated in binding buffer, and immunoprecipitated. Proteins were separated by SDS-PAGE and stained with Coomassie. S2: Strep-tag.

(F) FLAG-CWC22 and FLAG-BTZ were immunoprecipitated as described in (A).

(G) Immunoprecipitation of CWC22 and EJC proteins in vitro. Recombinant Strep-eIF4A3, MAGOH, Y14-Strep, BTZ (110–372)-Strep (lanes 1 and 2), or Flag-BTZ (110–372)-Strep (lane 3) were incubated under EJC-assembly conditions (see Experimental Procedures) in the presence (lane 2) or absence (lanes 1 and 3) of recombinant GST-CWC22 (110–409)-Flag. Immunoprecipitation was performed as in (E).

See also Figure S1.

database for spliceosomal proteins with homology to known eIF4A3-interacting proteins. Interestingly, we found that NOM1, a protein previously shown to interact with eIF4A3 during rRNA biogenesis in yeast and humans, shares significant sequence identity with the spliceosomal protein CWC22 (Figure S1). Notably, the highest sequence conservation between CWC22 and NOM1 was found within their MIF4G (*m*iddle domain of eukaryotic *i*nitiation *f*actor *4G* [eIF4G]) and MA3 domains. The MIF4G and MA3 domains of eIF4G have been shown to confer eIF4A binding during eukaryotic translation initiation (Marintchev et al., 2009), and likewise, mutations in the MIF4G domain of NOM1 defined its genetic interaction with eIF4A3 during pre-rRNA processing (Alexandrov et al., 2011).

To test whether CWC22 is indeed an eIF4A3-interacting factor, we immunoprecipitated FLAG-tagged CWC22 from HeLa cell lysates treated with RNase A (Figure 1A) or RNase I (Figure S1B) and analyzed the interaction with eIF4A3. Endogenous eIF4A3 coprecipitated with FLAG-CWC22 (Figures 1A and S1B), and this interaction was confirmed by the coprecipitation of endogenous CWC22 with FLAG-eIF4A3 (Figures 1B and S1C). Notably, two bands of coprecipitated CWC22 were visible, suggesting that posttranslational modifications may modulate its interaction with eIF4A3 (Figure 1B). The interaction of eIF4A3 with CWC22 is specific, because the RNA-binding proteins eIF4A1 and CBP80 fail to coprecipitate with CWC22 and eIF4A3, respectively (Figures 1A, 1B, S1B, and S1C).

The high degree of sequence conservation between the MIF4G and MA3 domains of NOM1 and CWC22 suggests that either one or both of these domains are involved in the binding of eIF4A3. To identify the eIF4A3-interacting domain(s) of CWC22, we performed coimmunoprecipitation experiments with N- and C-terminal truncations of FLAG-CWC22 (Figures 1C and 1D). The minimal eIF4A3-binding module identified consisted of the amino acid residues 110–409 of CWC22 (Figure 1D, lane 5). Notably, this part of CWC22 contains the MIF4G domain. In contrast, the MA3-domain-containing fragment (340–665) of CWC22 did not interact with eIF4A3 (Figure 1D, lane 6).

To analyze whether CWC22 interacts directly with eIF4A3, we purified full-length FLAG-Strep-tagged eIF4A3 and GST-tagged CWC22 from *E. coli.* We were not able to obtain the full-length

CWC22 protein from bacterial cultures, but three truncated versions, comprising residues 110–665, 110–409 and 340–665 of CWC22, were highly expressed. These recombinant CWC22 protein variants were preincubated with eIF4A3, and complex formation was analyzed by FLAG immunoprecipitation (Figure 1E). Both MIF4G-containing variants of CWC22 (110–665 and 110–409) but not the MA3 domain (340–665) readily copurified with eIF4A3, clearly demonstrating that the MIF4G domain of CWC22 directly interacts with eIF4A3 in vitro.

CWC22 Binds to eIF4A3 but Not MAGOH and Y14

Our data show that CWC22 directly binds eIF4A3, but whether this interaction involves the other core EJC proteins remains unsolved. To this end, we immunoprecipitated FLAG-BTZ (110–372) and FLAG-CWC22 from HeLa cell lysates and compared their interaction with the endogenous EJC core proteins eIF4A3, Y14, and MAGOH (Figure 1F). Both FLAG-BTZ (110–372) and FLAG-CWC22 copurified comparable amounts of endogenous eIF4A3. In contrast to FLAG-BTZ (110–372), FLAG-CWC22 did not coprecipitate the other EJC core proteins MAGOH and Y14 (Figure 1F, lane 2). Likewise, no interaction of BTZ with CWC22 was observed (Figure S1D). This indicates that CWC22 specifically binds to eIF4A3 outside of the assembled EJC.

This finding was confirmed in vitro with recombinant components (Figure 1G). The human EJC core was assembled from eIF4A3, MAGOH-Y14, and BTZ (110–372) as described previously (Ballut et al., 2005; Bono et al., 2006). Whereas FLAG-BTZ (110–372) coprecipitated all three core proteins of the EJC (lane 3), FLAG-CWC22 (110–409) coprecipitated only eIF4A3 and BTZ (lane 2). Although the short version of BTZ used here is coprecipitated together with eIF4A3 by CWC22, the full-length BTZ protein did not interact with CWC22 in cell lysates (Figure S1D). Hence, CWC22 recruits only eIF4A3 to the spliceosome, while the additional EJC components are presumably loaded by an alternative factor.

CWC22 Is Required for Pre-mRNA Splicing in Humans

We know that in yeast, Cwc22 is an essential splicing factor that is involved in the regulation of the first catalytic step via the DExD/H-box helicase Prp2 (Yeh et al., 2011). In contrast, the human CWC22 protein is only poorly characterized.

To assess whether human CWC22 is required for pre-mRNA splicing, we used two different small-interfering RNAs (siRNAs) to deplete CWC22 in human cell culture. In knockdown cells, CWC22 protein levels were reduced to ~20%, as shown by immunoblotting (Figure 2A). Northern blot analysis revealed that the levels of mature β-globin mRNA, expressed from an intron-containing reporter plasmid, were clearly reduced in CWC22-knockdown cells as compared to control cells (Figure 2B, lane 2; Figure S2C). At the same time, we observed a higher molecular weight band on the Northern blot at the size of the unspliced β-globin pre-mRNA. mRNA expressed from a triosephosphate isomerase (TPI) reporter (Figure 2C) and a T cell receptor (TCR) reporter (Figure S2D) showed a similar, but quantitatively more pronounced, effect. We used quantitative RT-PCR with intron- and exon-specific primer pairs (Figure S2A) to further analyze the β-globin and TPI mRNAs and pre-mRNAs from knockdown and control cells. In line with our Northern blot results, the levels of the pre-mRNA-specific PCR product were strongly increased, whereas the levels of the mRNA-specific PCR product were reduced in CWC22-knockdown cells (Figure 2D). To explore whether the CWC22 knockdown also impinges on endogenously expressed mRNAs, we quantified pre-mRNAs and mRNAs of two housekeeping genes, actin-beta (*ACTB*) and glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*). Notably, the primer pairs amplify different regions of pre-mRNA and mRNA, respectively, (Figure S2A) and therefore cannot discriminate between unspliced and partially spliced transcripts that may accumulate as a result of CWC22 depletion. However, decreased amounts of the spliced mRNA and increased amounts of the intron-containing pre-mRNA were observed for both genes in CWC22-depleted cells (Figure 2D).

The splicing defect of CWC22-depleted cells was rescued by the expression of siRNA-resistant CWC22 (Figure 2B, lane 3; Figure 2C, lane 3; Figure S2D, lane 3). This finding confirmed the specificity of the knockdown and enabled us to test which of the variants of CWC22 is sufficient to support splicing in HeLa cells. We expressed the N- and C-terminal truncations of CWC22 that were previously used to identify the binding domain of eIF4A3 in CWC22-knockdown cells. Interestingly, the CWC22 fragment containing both the MIF4G domain and adjacent MA3 domain (residues 110–665) was able to rescue the splicing defect. In contrast, the MIF4G domain (residues 110–409) or the MA3 domain (340–665) alone was insufficient to restore splicing in knockdown cells (Figure 2E), despite showing expression levels comparable to full-length CWC22 (Figure S2E). Together, our data demonstrate that the depletion of CWC22 strongly impairs pre-mRNA splicing in human cells and suggest a critical role for the MIF4G and MA3 domains during splicing.

The Interaction of eIF4A3 and CWC22 Displays Characteristics of an eIF4A-eIF4G Complex

So far, we have demonstrated that the MIF4G domains of CWC22 and eIF4A3 form a complex. Because of the high degree of sequence conservation between the DExD/H-box helicases eIF4A3 and eIF4A, the CWC22-eIF4A3 complex could resemble the well-characterized eIF4G-eIF4A complex. To generate a molecular model of the CWC22-eIF4A3 interaction, we superimposed the structures of eIF4A3 (Bono et al., 2006) and eIF4A (Schütz et al., 2008) and used the cocrystallized MIF4G of eIF4G as surrogate of CWC22 (Figure 3A). On the basis of this model, we generated mutants of CWC22 and eIF4A3 and tested them for their ability to form the CWC22-eIF4A3 complex.

We identified mutations (DLYD270KLYK, TI276GD, NFT301LAG), at three different positions in eIF4A3, that abolished the interaction with endogenous CWC22 in FLAG-immunoprecipitation assays (Figures 3B and S3A). The same mutations also abrogated the binding between recombinant eIF4A3 and CWC22 (110–409) (Figure 3C).

Likewise, the mutation of two amino acid residues at the predicted binding surface of CWC22 (NK171DE) impaired the coprecipitation of endogenous eIF4A3 from cell lysates (Figure 3D) as well as the interaction between recombinant eIF4A3 and CWC22 (Figures 3E and S3B). Interestingly, the mutation at position 171/172 of CWC22 maps to a 12-amino-acid-long

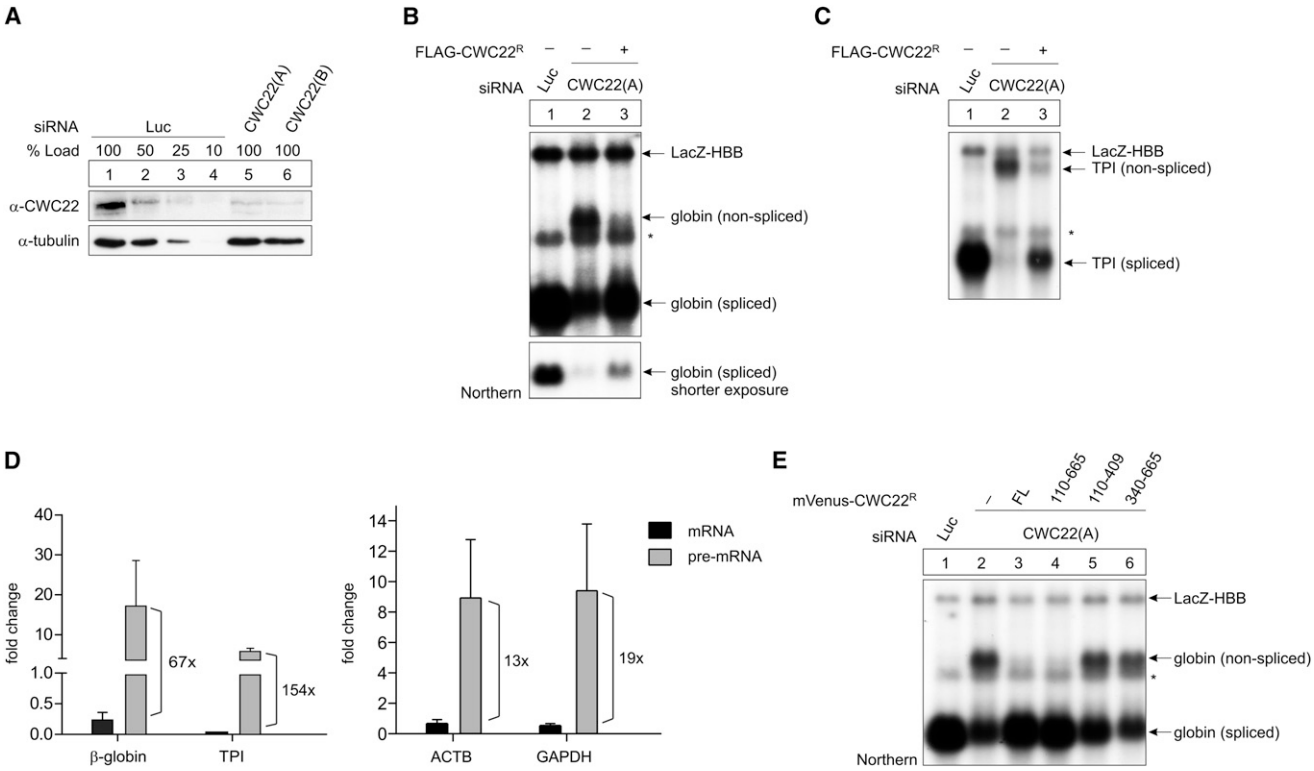


Figure 2. CWC22 Is an Essential Splicing Factor in Human Cells

(A) siRNA knockdown of CWC22 in human cell culture. HeLa cells were transfected with siRNAs targeting CWC22 (CWC[A] and CWC[B]) or Luciferase (negative control). The knockdown efficiency was assessed by immunoblotting with a CWC22-specific antibody. Tubulin served as a loading control.

(B, C, and E) Northern blot analysis of β-globin and TPI reporter mRNAs in CWC22-depleted cells.

(B and C) An intron-containing β-globin (B) or TPI (C) reporter was coexpressed with a transfection control (LacZ-HBB) and siRNA-resistant FLAG-CWC22^R (lane 3) or unfused FLAG (lanes 1 and 2) in CWC22-knockdown (lanes 2 and 3) and control cells (lane 1). The mRNA and pre-mRNA was detected by Northern blotting with a specific probe. An asterisk denotes an unspecific band. The bottom panel (B) shows a shorter exposure of the fully spliced β-globin band.

(D) Expression levels of pre-mRNA and mRNA were determined by quantitative RT-PCR. Bars represent the fold change in mRNA or pre-mRNA levels upon CWC22 depletion, and standard deviations are indicated by error bars. Numbers next to bars represent the calculated fold splicing inhibition (see Experimental Procedures).

(E) CWC22-knockdown cells were cotransfected with the indicated N- and C-terminal truncations of FLAG-CWC22^R and the β-globin reporter as described in (B). An asterisk denotes an unspecific band.

See also Figure S2.

conserved sequence motif in eIF4G that has been shown to compose the largest contiguous interface with eIF4A in yeast (Schütz et al., 2008). Moreover, the corresponding surface residues of the yeast NOM1 ortholog, Sgd1p, mediate its genetic interaction with yeast eIF4A3 (Fal1p) (Figure S1, asterisk) (Alexandrov et al., 2011). The identification of binding residues at the predicted eIF4A3-CWC22 interface thus corroborates the hypothesis that CWC22 and eIF4A3 constitute an eIF4A-eIF4G-like interaction pair.

Binding of eIF4A3 to CWC22 Is Required for Splicing-Dependent EJC Assembly

Our data revealed that CWC22 is an essential splicing factor in human cells and interacts with eIF4A3. Since EJC-assembly is a strictly splicing-dependent process, we hypothesized that CWC22 functionally bridges the spliceosome and the EJC.

To test the effect of CWC22 on EJC assembly in vitro, we incubated purified CWC22 (110–409) with equal amounts of recombinant EJC core components eIF4A3, BTZ, Y14, and MAGOH. In line with the results of our immunoprecipitations from cell lysates (Figure 1F), CWC22 did not coprecipitate together with FLAG-Y14 and the assembled EJC (Figure S4A, lane 3). Interestingly, the presence of wild-type CWC22, but not of the eIF4A3-interaction-deficient mutant of CWC22 (171DE), clearly reduced the amounts of coprecipitated eIF4A3 and BTZ (Figure S4A). Hence, the interaction of CWC22 with eIF4A3 impairs the assembly of recombinant EJCs in vitro. This suggests that additional spliceosomal factors or full-length CWC22 are required in order to recapitulate splicing-dependent EJC assembly in vitro.

Splicing-dependent EJC assembly has previously been studied by means of in vitro splicing reactions (Gehring et al., 2009a). Hence, we utilized splicing assays to test whether the CWC22-binding-deficient eIF4A3 mutants (DLYD270KLYK, TI276GD, NFT301LAG) are incorporated into EJCs. To this end, FLAG-containing mRNPs were immunoprecipitated,

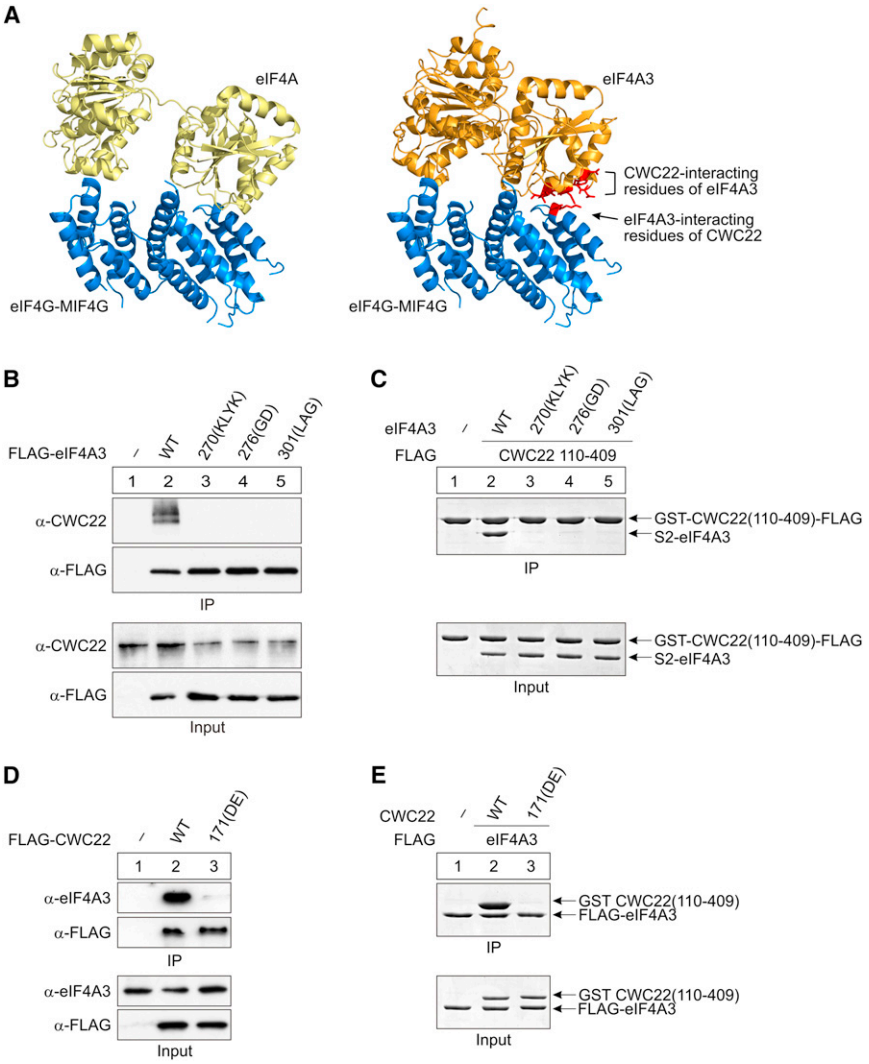


Figure 3. The Interaction of eIF4A3 and CWC22-MIF4G Features Characteristics of the eIF4A-eIF4G Complex

(A) The crystal structure of eIF4A and the MIF4G domain of eIF4G (left side, PDB 2VSO) was used to generate a model (right side) of the eIF4A3-CWC22 interaction by superposition of the structures of eIF4A3 (as part of the EJC; PDB 2JOS) and eIF4A. Putative contact sites (depicted in red) of eIF4A3 and CWC22 were mutated and characterized in (B–E). The superposition was generated by DaliLite (Holm and Park, 2000); the drawing was rendered with PyMOL (DeLano, 2010).

(B and C) Analysis of eIF4A3 mutant proteins. (B) FLAG-eIF4A3 and mutants of FLAG-eIF4A3 were immunoprecipitated from RNase-A-treated cell lysates. Coprecipitated endogenous CWC22 was detected by immunostaining with a specific antibody.

(C) Recombinant Strep-eIF4A3, Strep-eIF4A3 mutants, and GST-CWC22 (110–409)-FLAG were incubated in binding buffer and immunoprecipitated. Proteins were separated by SDS-PAGE and stained with Coomassie.

(D and E) Analysis of a CWC22 mutant protein. (D) FLAG-CWC22 and FLAG-CWC22 (NK171DE) were immunoprecipitated from RNase-A-treated cell lysates. Coprecipitated endogenous eIF4A3 was detected by immunostaining with a specific antibody.

(E) Recombinant FLAG-Strep-eIF4A3, GST-CWC22 (110–409) and GST-CWC22 (110–409/NK171DE) were incubated in binding buffer and immunoprecipitated as described in (C).

See also Figure S3.

followed by RNA extraction and denaturing PAGE. As expected, wild-type FLAG-eIF4A3 specifically coprecipitated spliced mRNA (Figure 4A, lane 2, top panel), demonstrating that it had been incorporated into an EJC (Gehring et al., 2009a). In contrast, none of the CWC22-binding-deficient variants of eIF4A3 precipitated any mRNA, indicating that the inability to interact with CWC22 prevents splicing-dependent EJC assembly (Figure 4A, lanes 3–5). Likewise, only the wild-type eIF4A3, but none of the mutants, precipitated splicing intermediates when the splicing reactions were stalled at the step of spliceosomal C-complex formation (MINX-GG; Figure 4A, bottom panel). Hence, CWC22 recruits eIF4A3 during the first catalytic step of splicing and thereby initiates EJC assembly.

The mutated residues in eIF4A3 (DLYD270KLYK, TI276GD, NFT301LAG) do not overlap with the interaction surfaces of Y14, MAGOH, or BTZ on eIF4A3 (Figure S4B). Nevertheless, we wanted to validate that the mutations do not disrupt the interaction of eIF4A3 with the other EJC components. To this end, we used recombinant proteins to assemble a splicing-independent EJC in vitro. Recombinant EJCs were assembled from the core proteins eIF4A3, Y14, MAGOH, and FLAG-BTZ (110–372) and immunoprecipitated via FLAG (Figure 4B). As expected, all core EJC proteins coprecipitated with FLAG-BTZ (110–372) (Figure 4B, lane 1). In contrast, no EJCs formed in the absence of ssRNA and AMP-PNP, and FLAG-BTZ (110–372) coprecipitated only eIF4A3 (Figure 4B, lane 5).

Notably, all three eIF4A3 mutants formed recombinant EJCs as efficiently as the wild-type (Figure 4B, lanes 2–4), clearly showing that the mutations affect not the ability to interact with the EJC core but, rather, the splicing-dependent loading of the complex. Hence, our results identify key residues of eIF4A3 that mediate both the interaction with the MIF4G domain of CWC22 and the assembly of EJCs by the spliceosome. Together with previously published data, our work thus redefines the framework of interactions required for the formation of EJCs in living cells.

DISCUSSION

Exon junction complexes are important functional mRNP components that assemble on maturing mRNA during splicing in the nucleus. In this work, we have identified the protein

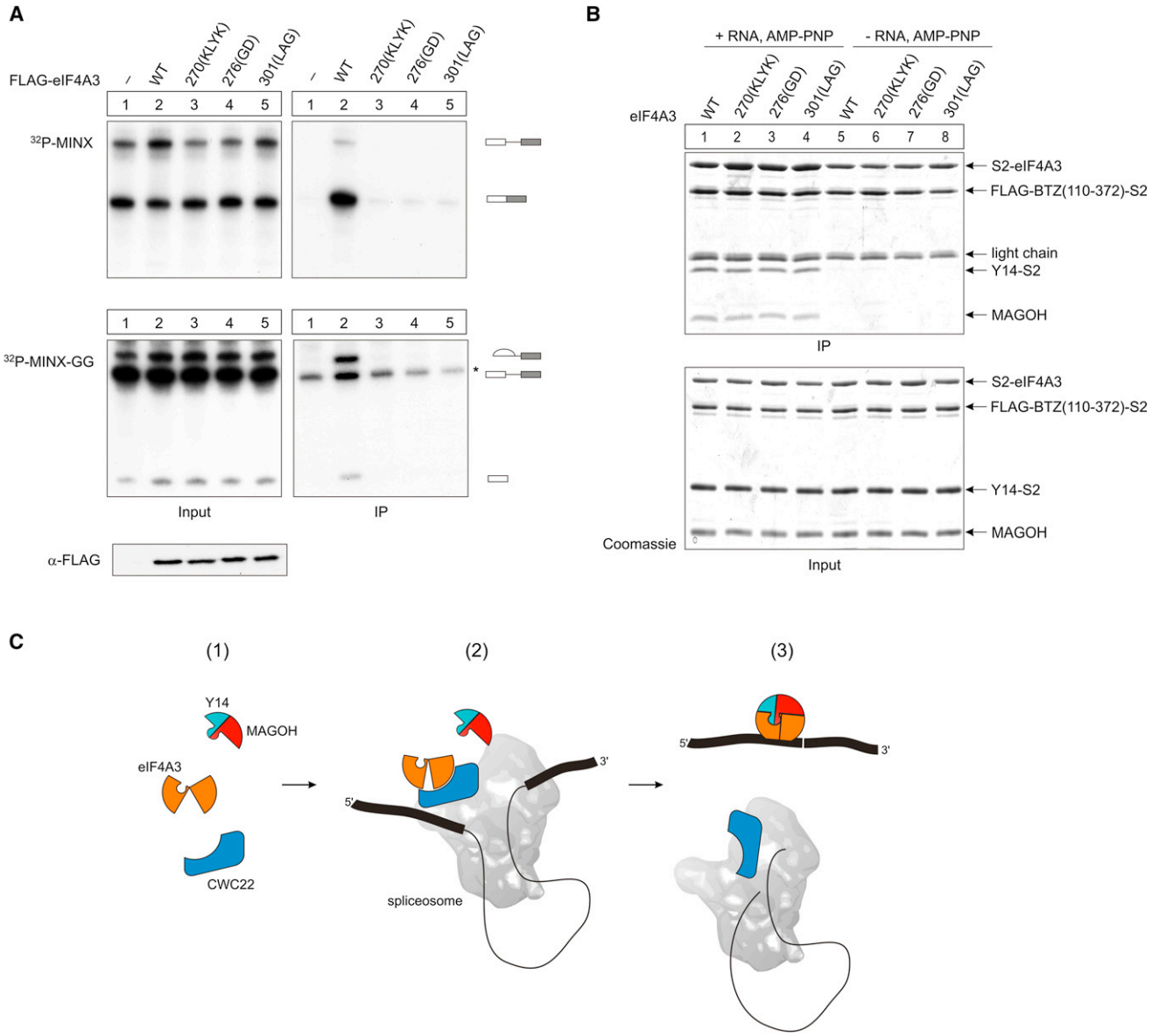


Figure 4. The Interaction of eIF4A3 and CWC22 Contributes to Splicing-Dependent EJC Assembly

(A) Analysis of splicing-dependent EJC assembly. Splicing reactions using MINX (top panel) or MINX-GG (mutated 3′ splice site, bottom panel) as substrate mRNA were carried out in splicing extracts composed of HeLa nuclear extracts and 293 whole-cell extracts expressing FLAG-eIF4A3, the indicated mutants of eIF4A3, or unfused FLAG. FLAG-containing mRNPs were immunoprecipitated, and coprecipitated RNA was separated by denaturing PAGE. Positions of unspliced precursor transcript, splicing intermediates (MINX-GG), and spliced product (MINX) are indicated schematically. Expression levels of the eIF4A3 mutants were determined by immunoblotting (α-FLAG, bottom).

(B) Analysis of splicing-independent EJC assembly in vitro from recombinant components. Strep-eIF4A3 (wild-type or the indicated mutants), FLAG-BTZ (110–372)-Strep, Y14-Strep, and MAGOH were incubated under EJC-assembly conditions in the presence (lanes 1–4) or absence (lanes 5–8) of AMP-PNP and ssRNA. Protein complexes were immunoprecipitated, separated by SDS-PAGE, and stained with Coomassie.

(C) The three-step model of splicing-dependent EJC deposition. For details, see Discussion.

See also Figure S4.

CWC22 as a central EJC-assembly factor. CWC22 is an abundant component of the activated spliceosome and the spliceosomal C complex (Bessonov et al., 2008). This supports the prior notion that eIF4A3 is recruited to the spliceosome during an early step of splicing (Gehring et al., 2009a; Mishler et al., 2008). Considering the molecular architecture of CWC22, its resemblance to other DExD-box-helicase-interacting proteins and its exclusive binding to eIF4A3, but not MAGOH and Y14, we propose a three-step model for EJC assembly (Figure 4C).

The binding of eIF4A3 to CWC22 represents the first step of EJC assembly (step 1). Although biochemically well defined, more details of this interaction need to be determined; in particular, whether CWC22 and eIF4A3 are recruited to the spliceosome individually or as a heterodimer. Previous studies showed that the binding of DExD-box helicases (e.g., eIF4A and the nuclear export factor Dbp5) to their interacting proteins (eIF4G and Gle1, respectively) stabilizes them in a "half-open" and activated conformation (Hilbert et al., 2011; Montpetit et al., 2011; Schütz et al., 2008). We therefore propose that eIF4A3 adopts a similar "half-open" conformation upon binding to CWC22 (step 2). This activated conformation would prevent premature RNA binding and prime eIF4A3 for receipt of the MAGOH-Y14 heterodimer. At the same time, the activated conformation of eIF4A3 enables it to be loaded to the mRNA in a single step through a conformational change imposed by MAGOH-Y14 binding (step 3). However, the molecular details of the MAGOH-Y14 recruitment to the spliceosome remain elusive and thus represent an interesting subject for future research.

The function of CWC22 within the spliceosome of mammalian cells is poorly understood. Although our data demonstrate that CWC22 is required for pre-mRNA splicing, its precise role in this process remains to be determined. Interactions between CWC22 and two spliceosomal proteins, FAM32A and FRG1, have been discovered in a large-scale Y2H-screen, but neither of these direct interaction partners elucidates the role of CWC22 in splicing (Hegele et al., 2012). Interestingly, the central part of CWC22 comprising the MIF4G and the MA3 domain, but not the MIF4G domain alone, is sufficient to rescue the splicing defect of CWC22-depleted cells. This suggests that an interaction involving the MA3 domain is required for the proper function of CWC22 during splicing. The yeast ortholog, Cwc22, has been described as an essential pre-mRNA splicing factor that is required for the function of Prp2 in promoting the release of SF3a and SF3b during the first catalytic step of splicing (Yeh et al., 2011). Similar to human CWC22, the MA3 domain of Cwc22 is required to rescue the splicing defect of Cwc22 knockout cells (Yeh et al., 2011). In order to shed light on the splicing function of CWC22, it will be instrumental to identify interaction partners of Cwc22 and CWC22 within the spliceosome that bind specifically to the MA3 domain.

Recently, a specific splicing function of the EJC core proteins eIF4A3, MAGOH, and Y14 has been described, but the precise role of the EJC during splicing remains obscure (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010). Considering the direct interaction between eIF4A3 and CWC22, it is tempting to speculate that the splicing function of the EJC is mediated via this molecular contact, although the deposition of EJCs is presumably not important for the general splicing function of CWC22.

Our description of the CWC22-eIF4A3 interaction provides a rationale for the intimate coupling of EJC deposition and pre-mRNA splicing. The conserved mode of interaction suggests that the activation and loading of eIF4A3 by CWC22 obeys the rules described for eIF4G, Gle1, and their DExD -box helicase interaction partners and proposes a general mechanism for DEAD-box helicase recruitment. Future studies will provide insight into the mechanistic and structural details of this interaction and reveal additional factors that are employed by the spliceosome to deposit the EJC.

EXPERIMENTAL PROCEDURES

Cell Culture, Plasmid Transfections, and siRNA Transfections

HeLa cells were grown and transfected with plasmid DNA or siRNA as previously described (Gehring et al., 2009a). For transfections of β-globin, TPI, and TCR, we used 1.5 μg reporter plasmid, 1.5 μg LacZ-HBB control plasmid, 0.5 μg rescue plasmid, and 0.3 μg GFP expression vector. Transfections for immunoprecipitations and 293 whole extracts were performed as previously described (Gehring et al., 2009a).

RNA Extraction and Analysis

Total RNA was extracted with TRIzol and analyzed by Northern blotting as previously described (Gehring et al., 2009a).

Immunoblot Analysis and Immunoprecipitation

FLAG-complexes were immunoprecipitated from RNase A (50 μg/ml)- or RNase I (25 U/ml)-treated HeLa cell lysates with M2 anti-FLAG magnetic beads (Sigma) at 4°C for 2 hr in lysis buffer (50 mM Tris [pH 7.2], 150 mM NaCl, 0.5% Triton X-100, + protease inhibitor). Beads were washed with lysis buffer, and complexes were eluted with SDS-sample buffer, separated by SDS-PAGE, and analyzed by immunoblotting.

In Vitro Transcription, In Vitro Splicing, and RNP Immunoprecipitation

Capped transcripts were generated by in vitro transcription with SP6 RNA polymerase in the presence of α^{32}P-GTP and cap analog (Promega). In vitro splicing reactions were performed in HeLa cell nuclear extract (CIL Biotech) that was supplemented with 293 whole-cell extracts (Gehring et al., 2009a). FLAG immunoprecipitations of RNPs were performed with FLAG-M2 affinity gel in mRNP IP buffer (20 mM HEPES KOH [pH 7.9], 200 mM NaCl, 2 mM $MgCl_2$, 0.2% Triton X-100, 0.1% NP-40, 0.05% Na-Deoxycholate). RNAs were recovered by TRIzol extraction and analyzed by denaturing PAGE.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.08.017.

LICENSING INFORMATION

ACKNOWLEDGMENTS

We thank Heidi Thelen and Juliane Hancke for excellent technical assistance; the CECAD Cologne imaging facility and the Uhlirova lab for microscopy advice; the Leptin and Uhlirova labs for sharing equipment; Fulvia Bono, Kusum Singh, and Benjamin Weiche for critical reading of the manuscript; and members of the Gehring lab for useful discussions. A.S. and V.B. are supported by fellowships from the International Graduate School in Development Health and Disease. This work was funded by a grant (GE2014/2-1) from the Deutsche Forschungsgemeinschaft to N.H.G.

Received: June 4, 2012
Revised: August 1, 2012
Accepted: August 23, 2012
Published online: September 6, 2012

REFERENCES

Alexandrov, A., Colognori, D., and Steitz, J.A. (2011). Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. Genes Dev. *25*, 1078–1090.

Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Séraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science *313*, 1968–1972.

Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrançois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P., and Therrien, M. (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. Cell *143*, 251–262.

Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Séraphin, B., and Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. Nat. Struct. Mol. Biol. *12*, 861–869.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846–850.

Bono, F., and Gehring, N.H. (2011). Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA Biol. *8*, 24–29.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell *126*, 713–725.

DeLano, W.L. (2010). The PyMOL Molecular Graphics System, Version 1.3r1 (New York: Schrodinger, LLC).

Gehring, N.H., Lamprinaki, S., Hentze, M.W., and Kulozik, A.E. (2009a). The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS Biol. *7*, e1000120.

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009b). Disassembly of exon junction complexes by PYM. Cell *137*, 536–548.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol. Cell *45*, 567–580.

Hilbert, M., Kebbel, F., Gubaev, A., and Klostermeier, D. (2011). eIF4G stimulates the activity of the DEAD box protein eIF4A by a conformational guidance mechanism. Nucleic Acids Res. *39*, 2260–2270.

Holm, L., and Park, J. (2000). DaliLite workbench for protein structure comparison. Bioinformatics *16*, 566–567.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J. *19*, 6860–6869.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J. *20*, 4987–4997.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. Science *293*, 1836–1839.

Ma, X.M., Yoon, S.O., Richardson, C.J., Jülich, K., and Blenis, J. (2008). SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. Cell *133*, 303–313.

Marintchev, A., Edmonds, K.A., Marintcheva, B., Hendrickson, E., Oberer, M., Suzuki, C., Herdy, B., Sonenberg, N., and Wagner, G. (2009). Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. Cell *136*, 447–460.

Mishler, D.M., Christ, A.B., and Steitz, J.A. (2008). Flexibility in the site of exon junction complex deposition revealed by functional group and RNA secondary structure alterations in the splicing substrate. RNA *14*, 2657–2670.

Montpetit, B., Thomsen, N.D., Helmke, K.J., Seeliger, M.A., Berger, J.M., and Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. Nature *472*, 238–242.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688–700.

Rodríguez-Navarro, S., and Hurt, E. (2011). Linking gene regulation to mRNA production and export. Curr. Opin. Cell Biol. *23*, 302–309.

Roignant, J.Y., and Treisman, J.E. (2010). Exon junction complex subunits are required to splice Drosophila MAP kinase, a large heterochromatic gene. Cell *143*, 238–250.

Schütz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. Proc. Natl. Acad. Sci. USA *105*, 9564–9569.

Tange, T.O., Shibuya, T., Jurica, M.S., and Moore, M.J. (2005). Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. RNA *11*, 1869–1883.

Yeh, T.C., Liu, H.L., Chung, C.S., Wu, N.Y., Liu, Y.C., and Cheng, S.C. (2011). Splicing factor Cwc22 is required for the function of Prp2 and for the spliceosome to escape from a futile pathway. Mol. Cell. Biol. *31*, 43–53.

Cell PRESS
Open ACCESS

EXTENDED EXPERIMENTAL PROCEDURES

Plasmids

Plasmid constructs MINX, MINX-GG, β-globin wt, pCI-FLAG, pCI-V5, and expression vectors for eIF4A3 and BTZ(110-372) were described previously (Gehring et al., 2009a; Gehring et al., 2009b). TPI (synthetic gene, Geneart) and TCR (kindly provided by Miles Wilkinson) reporter constructs were cloned into pCI-neo and fused at their 3′end to a β-globin 3′UTR sequence to enable detection by a β-globin probe. The LacZ-HBB control plasmid was generated by inserting a β-globin 3′UTR sequence to pcDNA4/TO/lacZ. Full length CWC22 cDNA was obtained by RT-PCR using total HeLa cell RNA and inserted into pCI-neo-FLAG. Mutants of eIF4A3 and CWC22 were generated by site-directed mutagenesis. All constructs were verified by DNA sequencing.

siRNA Transfection

The siRNA target sequences for CWC22 are 5′-AAAGTAGTGTGGCACAGATAA-3′ (siRNA CWC[A]) and 5′-CTCGCACTGGTGGAG CATATA-3′ (siRNA CWC[B]). For the complementation of the CWC22 depletion 0.5 μg of a siRNA insensitive CWC22 expression plasmid was transfected together with the reporter plasmid. This construct was generated by replacing the siRNA CWC(A) targeting sequence 5′-AAA AGT AGT GTG GCA CAG ATA AAA-3′ (codons 2-9 of the CWC22 ORF) by the resistant sequence 5′-AAA TCA TCA GTG GCC CAA ATC AAA-3′.

Antibodies

The antibodies against CWC22, Y14, FLAG and V5 were from Sigma, the antibodies against MAGOH and eIF4A1 were from Abcam. The eIF4A3 polyclonal antibody was made by GenScript with an N-terminal peptide of eIF4A3. The CBP80 antibody was kindly provided by Elisa Izaurralde.

Protein Preparation

Codon-optimized eIF4A3 with an N-terminal Strep-tag was inserted into pET51 (Novagen). Y14 with a C-terminal Strep-tag and MAGOH were cloned into MCS1 and MCS2 of pETDuet-1 (Novagen), respectively. The SELOR domain of BTZ (residues 110-372) with a C-terminal Strep-tag, with or without an N-terminal FLAG-tag was inserted into pETDuet-1. Three fragments of CWC22 containing the MIF4G domain (residues 110-665, 110-409 or 340-665) were cloned into pGEX6P3 (GE Healthcare).

Recombinant proteins were expressed in *E.coli* BL21 star or Rosetta II pLysS over night at 28°C in autoinduction medium (Studier, 2005). Strep-eIF4A3, MAGOH/ Y14-Strep, FLAG-BTZ (110-372)-Strep and BTZ (110-372)-Strep were purified on a StrepTactin Superflow Plus Cartridge (QIAGEN). GST-CWC22 (110-409), GST-CWC22 (110-665) and GST-CWC22 (340-665) were purified on a GSTrap FF (GE Healthcare) column. Proteins were concentrated by ultrafiltration and stored in EJC binding buffer (20mM HEPES KOH (pH7.9), 125mM NaCl, 1mM $MgCl_2$, 2mM $CaCl_2$, 1mM DTT, 2.5% Glycerol, 0.1% NP-40).

Binding Assays with Recombinant Proteins

For in vitro EJC assembly, recombinant proteins (15μg each), 0.5μmole AMP-PNP and 1nmole poly(U)$_{15}$ ssRNA where mixed in a total volume of 60μl EJC binding buffer and incubated over night at 4°C. FLAG-immunoprecipitations were performed with FLAG-M2 magnetic beads (Sigma) in EJC binding buffer for 2 hr at 4°C. Precipitated complexes were eluted with SDS-sample buffer, separated by SDS-PAGE and visualized by Coomassie staining. When indicated, binding was performed in the absence of AMP-PNP and ssRNA to prevent EJC formation.

For in vitro interaction studies, recombinant proteins (15μg each) were mixed in EJC binding buffer, preincubated 30 min on ice and immunoprecipitated with FLAG-M2 magnetic beads as described above.

Quantitative RT-PCR

Quantitative RT-PCR measuring SYBR Green incorporation was used to quantify pre-mRNA and mRNA expression levels of transfected reporters (β-globin and TPI reporter constructs) and endogenous genes (β-actin (ACTB) and GAPDH). Expression of reporter pre-mRNA and mRNA were normalized to the intronless APH 3′ II (neomycin resistance gene) expressed from the pCI-neo plasmid. To calculate the fold splicing inhibition, the pre-mRNA increase was divided by the mRNA reduction. Primer sequences were:

5′-AAGGCTCATGGCAAGAAAG-3′ and 5′-ACACCAGCCACCACTTTCT-3′ (β-globin mRNA)
5′-AGTCCAAGCTAGGCCCTTT-3′ and 5′-ACACCAGCCACCACTTTCT-3′ (β-globin pre-mRNA)
5′-AGTTCTTCGTTGGGGGAAAC-3′ and 5′-CCACAGCAATCTTGGGATCT-3′ (TPI mRNA)
5′-CTGGAAGGCTCTTCGAGTTG-3′ and 5′-CCACAGCAATCTTGGGATCT-3′ (TPI pre-mRNA)
5′-ATACTTTCTCGGCAGGAGCA-3′ and 5′-TGAATGAACTGCAGGACGAG-3′ (APH 3′ II)
5′-GGACTTCGAGCAAGAGATGG-3′ and 5′-AGCACTGTGTTGGCGTACAG-3′ (ACTB mRNA)
5′- AGAAAATCTGGCACCACACC-3′ and 5′-AACGGCAGAAGAGAGAACCA-3′ (ACTB pre-mRNA)
5′-GAGTCAACGGATTTGGTCGT-3′ and 5′-TTGATTTTGGAGGGATCTCG-3′ (GAPDH mRNA)
5′-GAGCTGGGGAATGGGACT-3′ and 5′-TGATGGCATGGACTGTGG-3′ (GAPDH pre-mRNA)

Microscopy

HeLa cells were transfected with 100 ng of pCI-mVenus and 100 ng pCI-Kate2 plasmids. Cells were plated on coverslips 24 hr post transfection. 48 hr after transfection cells were fixed with 3.7% formaldehyde. Images were taken using FV1000 confocal microscope (Olympus).

Protein-Complex Purification on Strep-Tactin Spin Columns

Proteins (15μg each) were pre-incubated for 30 min on ice and purified with the Strep-Tactin SpinPrep kit (Novagen) according to the manufacturer's instructions. Eluted proteins were separated by SDS-PAGE and stained with Coomassie.

SUPPLEMENTAL REFERENCES

Gehring, N.H., Lamprinaki, S., Hentze, M.W., and Kulozik, A.E. (2009a). The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS Biol. *7*, e1000120.

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009b). Disassembly of exon junction complexes by PYM. Cell *137*, 536–548.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

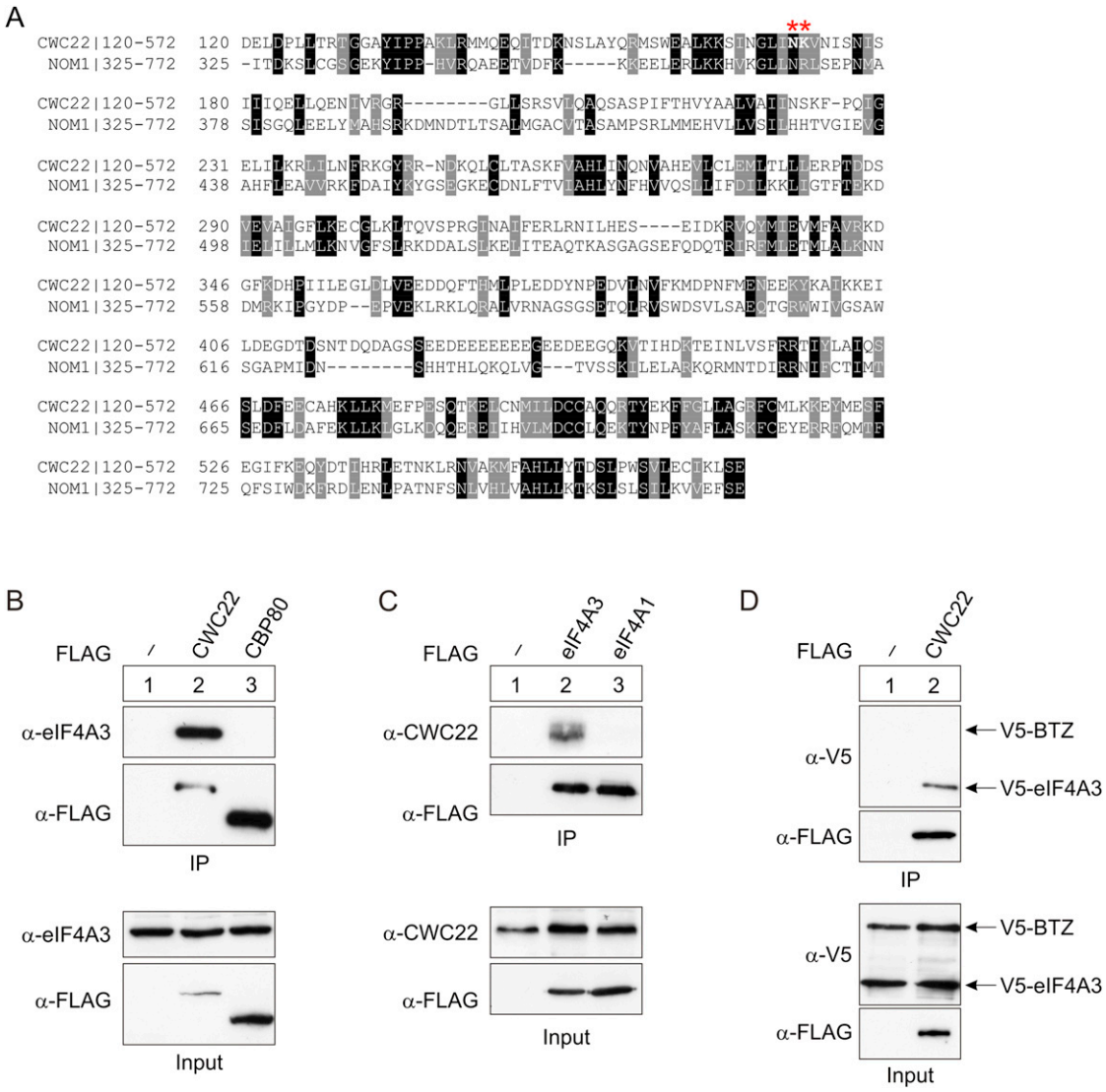


Figure S1. CWC22 Interacts with eIF4A3, Related to Figure 1

(A) Alignment of the protein sequences of human CWC22 (Q9HCG8) and NOM1 (Q5C9Z4). The alignment was generated using ClustalW and boxshade 3.21. The red asterisks indicate the eIF4A3-binding residues at position 171/172 of CWC22. (B) FLAG-CWC22 and FLAG-CBP80 were immunoprecipitated from RNase I treated cell lysates. Co-precipitated endogenous eIF4A3 was detected by immunostaining with a specific antibody. Unfused FLAG served as negative control. (C) FLAG-eIF4A3 and FLAG-eIF4A1 were immunoprecipitated as described in (B) and co-precipitated endogenous CWC22 detected by immunostaining with a specific antibody. Unfused FLAG served as a negative control. (D) HeLa cells were transfected with plasmids expressing V5-eIF4A3 and V5-BTZ together with FLAG-CWC22 (lane 2) or unfused FLAG (lane 1, negative control). Immunoprecipitation was performed as described in (B) and precipitated proteins were detected by immunostaining with the indicated antibodies.

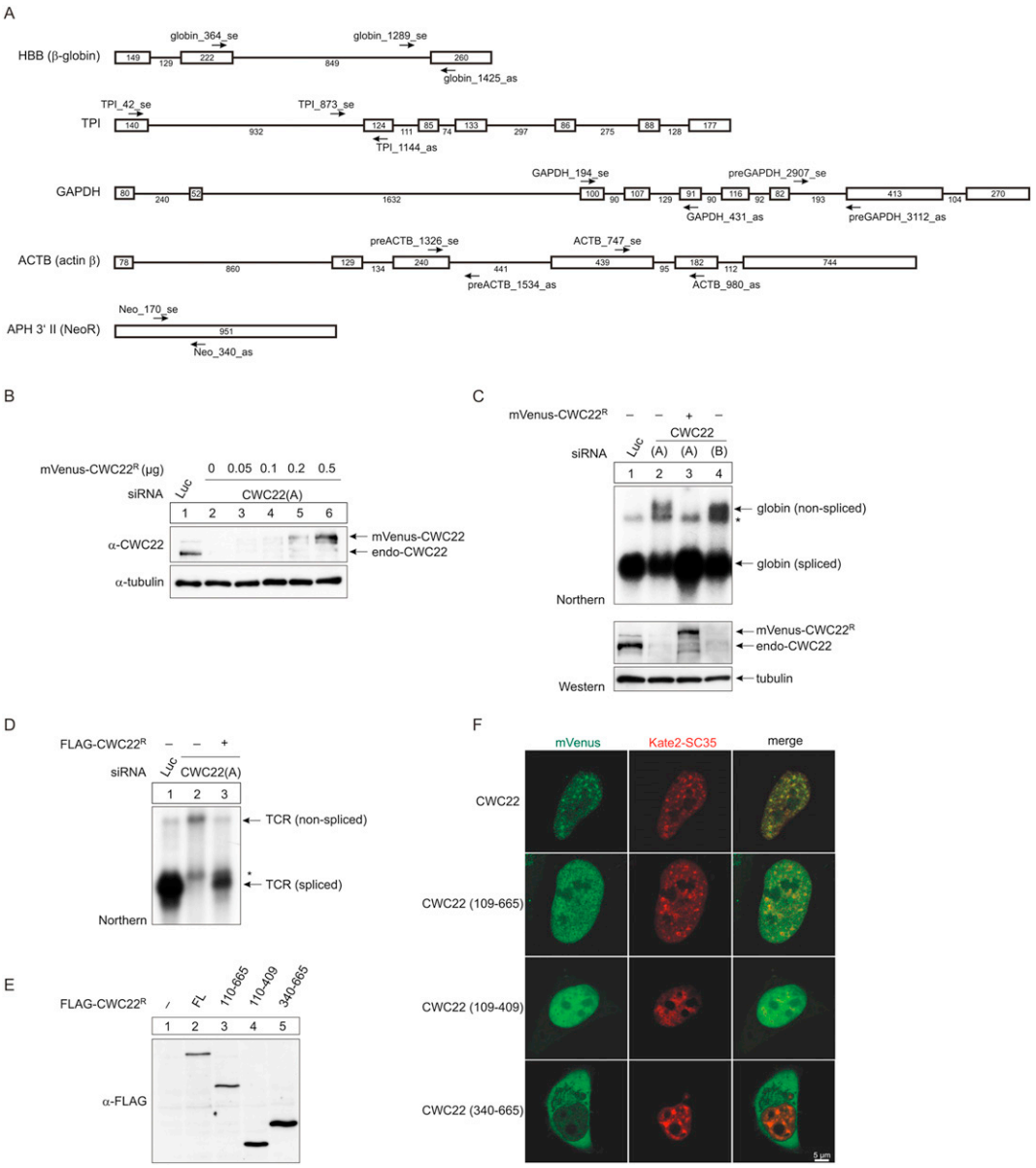


Figure S2. Functional Analysis and Subcellular Localization of CWC22, Related to Figure 2

(A) Exon-intron structure of intron-containing reporter genes (β-globin, TPI), endogenous genes (ACTB, GAPDH) and the intronless neomycin-resistance gene (APH 3′ II) that were analyzed by quantitative RT-PCR (see Figure 2D). Arrows denote the binding sites of mRNA- and pre-mRNA-specific primers. Numbers represent the lengths (in base pairs) of exons and introns.

(B) Expression levels of siRNA-resistant mVenus-CWC22^R-constructs. CWC22-knockdown cells were co-transfected with the indicated concentrations of plasmid expressing siRNA-resistant mVenus-CWC22^R. Expression levels were analyzed by immunostaining with anti-CWC22-antibody. Luciferase siRNA was used as a negative control; tubulin served as a loading control.

(C, D) Northern blot analyses of β-globin and TCR reporter mRNAs in CWC22-depleted cells.

(C) Cells were co-transfected with the β-globin reporter, siRNAs targeting Luciferase (lane 1) or CWC22 (lane 2,3 (siRNA[A]) and 4 (siRNA[B]) and mVenus-CWC22^R (lane 3) or unfused FLAG (lane 1, 2, 4). The mRNA and pre-mRNA was detected by Northern blotting with a specific probe. The asterisk denotes an unspecific band. Bottom panel: Expression of CWC22 protein in knockdown and control cells was analyzed by immunostaining with a specific antibody.

(D) Cells were co-transfected with the TCR reporter, siRNAs targeting Luciferase (lane 1) or CWC22 (lane 2,3, siRNA[A]) and FLAG-CWC22^R (lane 3) or unfused FLAG (lane 1, 2). Northern blot analysis was performed as described in (C). The asterisk denotes an unspecific band.

(E) Expression levels of the indicated CWC22^R-rescue constructs were analyzed by immunostaining with anti-FLAG-antibody.

(F) Subcellular localization of mVenus-CWC22 constructs. HeLa cells were grown on coverslips and transfected with plasmids expressing the indicated mVenus-CWC22-constructs and Kate2-SC35 as a marker for nuclear speckles. Cells were fixed 48hrs post-transfection and analyzed by confocal microscopy. Images of representative cells from each transfection were selected. Scale bar = 5μm. Full-lengths mVenus-CWC22 co-localizes with Kate2-SC35 in nuclear speckles whereas mVenus-CWC22 (110-665) and mVenus-CWC22 (110-409) display a diffuse nuclear pattern. mVenus-CWC22 (340-665) is cytoplasmic.

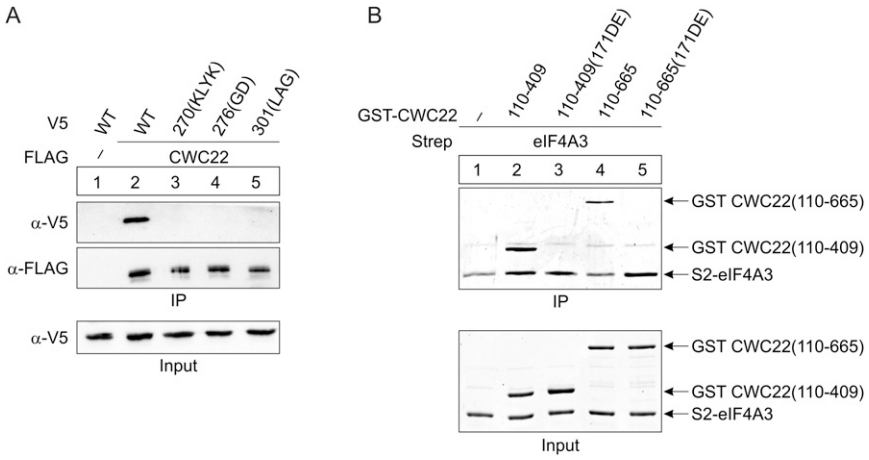


Figure S3. The Interaction of eIF4A3 and CWC22-MIF4G Features Characteristics of the eIF4A-eIF4G Complex, Related to Figure 3

(A) FLAG-immunoprecipitation of HeLa cell lysates expressing FLAG-tagged CWC22 or unfused FLAG (negative control) together with V5-tagged eIF4A3 (wt and indicated mutants). Co-precipitation of V5-eIF4A3 was shown by immunostaining with anti-V5 antibody.

(B) Strep-pulldown of recombinant proteins. Strep-eIF4A3 and indicated GST-CWC22-versions were purified from *E.coli,* incubated in binding buffer and purified on StrepTactin spin columns. Purified proteins were separated by SDS-PAGE and stained with Coomassie.

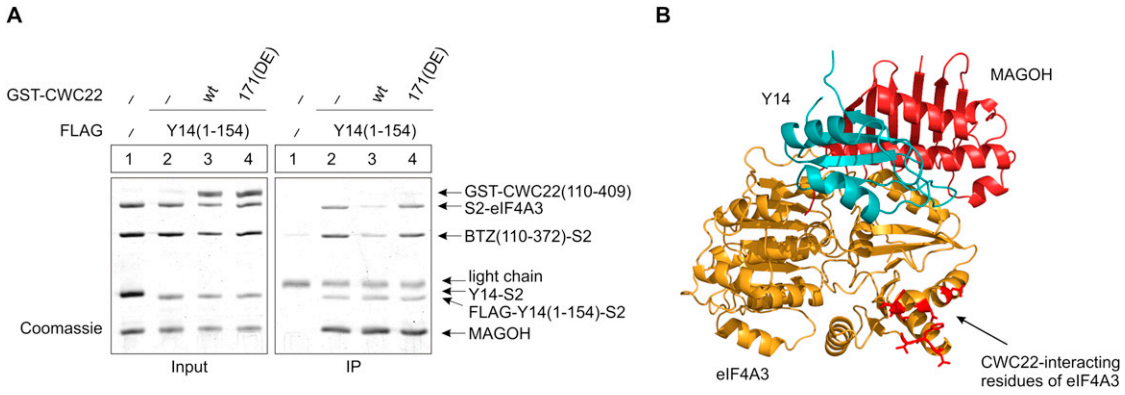


Figure S4. CWC22 Is Not Required for In Vitro EJC Assembly, Related to Figure 4

(A) CWC22 impairs EJC assembly in vitro. Recombinant Strep-eIF4A3, BTZ (110-372)-Strep, MAGOH and Y14-Strep (lane 1) or FLAG-Y14 (1-154)-Strep (lanes 2, 3, 4) were incubated under EJC assembly conditions in the presence of GST-CWC22 (110-409) (lane 3) or GST-CWC22 (110-409/171DE) (lane 4). Protein complexes were immunoprecipitated via FLAG. Precipitated proteins were separated by SDS-PAGE and stained with Coomassie.

(B) The mutated residues in eIF4A3 do not overlap with the interaction surfaces of Y14 or MAGOH. X-ray structure of the core EJC (PDB ID code 2JOS). Depicted are eIF4A3 (orange), Y14 (cyan) and MAGOH (red). The residues of eIF4A3 (270-274(DLYD), 276/277(TI), 301-303(NFT)) that are required for CWC22-binding are highlighted in red. BTZ binds at the opposite side of the eIF4A3 molecule and is not shown.

Published online 23 November 2013 Nucleic Acids Research, 2014, Vol. 42, No. 4 2673–2686
doi:10.1093/nar/gkt1197

Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2

Marcello Clerici[1,2], Aurélien Deniaud[1,2], Volker Boehm[3], Niels H. Gehring[3], Christiane Schaffitzel[1,2,*] and Stephen Cusack[1,2,*]

[1]European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France, [2]Unit of Virus Host-Cell Interactions, University of Grenoble Alpes-EMBL-CNRS, UMI 3265, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France and [3]University of Cologne, Institute for Genetics, Zuelpicher Street 47a, 50674 Cologne, Germany

Received July 16, 2013; Revised November 2, 2013; Accepted November 4, 2013

ABSTRACT

Nonsense-mediated decay (NMD) is a eukaryotic quality control pathway, involving conserved proteins UPF1, UPF2 and UPF3b, which detects and degrades mRNAs with premature stop codons. Human UPF2 comprises three tandem MIF4G domains and a C-terminal UPF1 binding region. MIF4G-3 binds UPF3b, but the specific functions of MIF4G-1 and MIF4G-2 are unknown. Crystal structures show that both MIF4G-1 and MIF4G-2 contain N-terminal capping helices essential for stabilization of the 10-helix MIF4G core and that MIF4G-2 interacts with MIF4G-3, forming a rigid assembly. The UPF2/UPF3b/SMG1 complex is thought to activate the kinase SMG1 to phosphorylate UPF1 *in vivo*. We identify MIF4G-3 as the binding site and *in vitro* substrate of SMG1 kinase and show that a ternary UPF2 MIF4G-3/UPF3b/SMG1 complex can form *in vitro*. Whereas *in vivo* complementation assays show that MIF4G-1 and MIF4G-2 are essential for NMD, tethering assays reveal that UPF2 truncated to only MIF4G-3 and the UPF1-binding region can still partially accomplish NMD. Thus UPF2 MIF4G-1 and MIF4G-2 appear to have a crucial scaffolding role, while MIF4G-3 is the key module required for triggering NMD.

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is a eukaryotic mRNA quality control mechanism that recognizes transcripts with premature termination codons (PTC) and promotes their degradation. Thereby, NMD protects the cell from the potentially deleterious effects of C-terminally truncated proteins (1–6). In fact, ~30% of all known disease-causing mutations in humans involve production of PTC-containing mRNAs (7). NMD also contributes to regulate the abundance of several physiological substrates, targeting 3–10% of the transcriptome in different organisms (8–10).

Nine NMD protein factors, SMG1-9, have been identified in most metazoans (11,12), and recently four additional factors (SMG10, RUVLB1, RUVBL2 and RPB5) have been added to the components of the NMD machinery (13). The three UPF (UP-Frameshift) proteins, UPF1 (SMG2), UPF2 (SMG3) and UPF3 (SMG4) constitute the conserved core of NMD and are found in almost all eukaryotes with a few possible exceptions among protists (14–16), suggesting an ancient evolutionary origin for NMD. UPF1 is an ATP-dependent RNA helicase that is directly involved in the recognition of terminating ribosomes stalled at a PTC (17–19). Human UPF2 (Q9HAU5) is an ~140 kDa protein containing three conserved MIF4G (Middle portion of eIF4G) domains that are found in a number of proteins involved in RNA metabolism and translation such as the nuclear cap-binding protein CBP80 and eIF4G (eukaryotic initiation factor 4-gamma) (20). UPF2 interacts via its third MIF4G domain with UPF3b and via its C-terminal extremity with UPF1, thus forming the central component of the ternary complex of UPF proteins (21–23). UPF3b stably binds the exon junction complex (EJC) (24–26), which is deposited by the splicing machinery on the mRNA ~24 nt upstream of the exon boundaries. The EJC functions as an enhancer of NMD

*To whom correspondence should be addressed. Tel: +33 476207238; Fax: +33 476207199; Email: cusack@embl.fr
Correspondence may also be addressed to Christiane Schaffitzel. Tel: +33 476207505; Fax: +33 476207199; Email: bschaff@embl.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

efficiency in mammals (27–29) probably by serving as a recruitment platform for UPF3b (22,24,26) and UPF2 after mRNA export to the cytoplasm (16,30).

Ribosomes stalled at a PTC are recognized by the NMD factors UPF1 and SMG1 to form the transient SURF complex, which consists of SMG1-UPF1-eRF1/3a, SMG8 and SMG9 (12,19). SMG1 is an ~415-kDa serine/threonine-protein kinase, essential for NMD in human and *Caenorhabditis elegans* (31), that belongs to the phosphatidylinositol 3-kinase–related kinase (PIKK) protein family (32). The EM structure of SMG1 in complex with SMG9 has been reported at 24 Å showing a characteristic head and arm architecture as observed previously for DNA-PKcs, another member of the PIKK family (33). Interaction of the SURF complex with NMD factors UPF2 and UPF3b positioned on a 3′ EJC is suggested to activate SMG1 to phosphorylate UPF1 in the 'decay-inducing complex' (DECID). UPF1 phosphorylation by SMG1 in the DECID leads to translation termination and dissociation of eRF3a from UPF1 (12,19). In addition, NMD factors SMG5-7 are recruited, which promote mRNA decay (34,35) and ultimately dephosphorylation of UPF1 (36). A major role in the interaction between the SURF and UPF2/3b-EJC complexes is played by UPF2 binding to UPF1. This interaction is mediated by the intrinsically disordered C-terminal extremity of UPF2, which structures on binding to the CH-rich domain of UPF1 (15). Moreover, an interaction between UPF2 and the SMG1 C-terminal region containing the kinase domain has been reported (19).

A recent cryo-EM structure of the UPF1/2/3-EJC complex indicates that UPF2 acts as a central scaffolding protein with a ring-like arrangement of the MIF4G domains (37). According to the quasi-atomic model, UPF2 forms the crucial contacts with UPF1, UPF3b and the EJC and positions UPF1 toward the 3′-end of the mRNA where it could exert its helicase activity during mRNA degradation (37). High-resolution structural information of UPF2 is limited to the structure of the MIF4G-3 domain in complex with the RNP domain of UPF3b (16) and the UPF1-binding region in complex with UPF1 (15). The latter is separated from MIF4G-3 by a conserved low-complexity acidic region (88 residues, ~50% Asp/Glu, indicated as LR3 in Figure 1A), which is likely to be disordered (15,23,38). Neither the structure nor the particular roles of the first two MIF4G domains of UPF2 are known. MIF4G domains have a conserved 3D structure, comprising an elongated bundle of antiparallel helices, despite high sequence divergence. They are thought to serve as molecular platforms for the recruitment of interaction partners (20,39–42) but could also be molecular spacers or scaffolds to correctly position distant functional regions.

Here, we present the atomic resolution crystal structures of the N-terminal (MIF4G-1) and middle (MIF4G-2) domains of UPF2 and a lower resolution crystal structure of the combined UPF2 MIF4G-2/MIF4G-3 domains. These structures allow a refinement of the quasi-atomic model of the entire UPF-EJC complex. Furthermore, we test the importance of UPF2 MIF4G-1 and MIF4G-2 domains for NMD *in vivo* and identify the UPF2 MIF4G-3 domain as the SMG1 interaction site. We also show that SMG1 kinase phosphorylates UPF2 MIF4G-3 domain *in vitro,* mainly at the serine residue 1046. *In vivo*, however, the UPF2 mutation S1046A did not interfere with NMD in complementation experiments. These results highlight the central role of the UPF2 MIF4G-3 domain in the DECID complex since it interacts with both UPF3 and SMG1 kinase, perhaps activating the latter to phosphorylate UPF1.

MATERIALS AND METHODS

Full methods are given in the Supplementary Material.

Crystallographic data collection and structure determination

Crystallographic statistics are given in Supplementary Table S1. All data collection was performed at 100 K at the European Synchrotron Radiation Facility. The data were integrated with XDS (43) and analyzed with CCP4i. Molecular replacement was performed with PHASER (44), model building with COOT (45) and refinement with REFMAC5 (46,47). The structures of both MIF4G-1 and MIF4G-2 were solved *de novo* by selenomethionine (SeMet) Single Anomalous Dispersion (SAD), using SHELXD (48) to find sites and SHARP (49) for refinement and phasing. For the low-resolution combined MIF4G-2/MIF4G-3 domain structure, molecular replacement using PHASER (44) using the individual domains, gave an unambiguous unique solution with log likelihood gain of 220. This was confirmed by correspondence of predicted methionine positions with anomalous difference peaks using data from MIF4G-2/MIF4G-3 crystals grown with SeMet. Due to the low resolution, no further refinement was performed.

Generation of the quasi-atomic UPF-EJC model

Starting from the EM reconstruction of the UPF-EJC complex (EMD-2048) and the corresponding quasi-atomic model [Melero *et al.* (37)], we replaced the UPF2 MIF4G domains and the UPF3b RNA-recognition motif (RRM) domain by our crystal structure of MIF4G-1 domain and by the MIG4G-2/3/UPF3b RRM complex. We used Chimera (50) to obtain the best correlation coefficient for the placement of the domains into the density. To avoid a clash with MIF4G-1 and MIF4G-2 (both are larger than the MIF4G homology model used in the original atomic model), the EJC had to be moved by 14 Å away from UPF2. Similarly, the UPF1 CH-domain was moved by 5 Å away from UPF2 MIF4G-3 (Supplementary Figure S7). The figures were generated with PyMOL (DeLano Scientific). In the rendering of the cryo-EM map, the density cutoff was set for the display of the envelope to represent ~130% of the *a priori* estimated volume.

In vivo NMD assays

UPF2 tethering assay with the β-globin 4MS2 construct and the transfection control (wt300+e3) were described previously (24,51). For NMD rescue experiments, HeLa

cells were transiently transfected with 100 pmol siRNA targeting UPF2 or luciferase. On the next day, cells were split 1:2 into 10-cm plates and 24 h later transfected with 200 pmol siRNA. The next day, cells were transfected by calcium phosphate precipitation with GFP and siRNA resistant UPF2 expression plasmids and with the plasmid encoding the reporter mRNA (TPI-HBB) (52).

Surface plasmon resonance

Surface plasmon resonance (SPR) experiments were performed on a BIAcore 3000 using SA sensor chips (GE-Healthcare). One flow cell was not functionalized to be used as a background control. The second flow cell was functionalized with purified SMG1 to a density of ~2000 RU. UPF2/UPF3b solutions were injected at 25 µl/min during 3 or 4 min followed by a 10-min dissociation phase. The surfaces were regenerated by 1-min injection of 0.5 M and 1 M NaCl solutions. The data were analyzed with the Biacore evaluation software by subtracting both the control flow cell and the buffer injection curve. Apparent equilibrium dissociation constant ($K_{D\text{-app}}$) were determined using the RU values measured 10 s before the end of the association phase for all curves (RU_{max}). These RU_{max} were plotted as a function of protein concentration and fitted assuming a one binding site model.

Data deposition

Coordinates and structure factors for the UPF2 MIF4G-1 and MIF4G-2 domains are deposited in the Protein Data Bank with accession codes 4CEM and 4CEK, respectively.

RESULTS

For structural studies of human UPF2 we initially used a construct comprising residues 121–1031 that encompasses all three predicted MIF4G domains but excludes the N-terminal low-complexity region (indicated as LR1 in Figure 1A) and the C-terminal acidic region as well as UPF1 binding domain (U1BD) (Figure 1A). UPF2(121–1031) was expressed in *Escherichia coli*, purified and subjected to crystallization trials, but no hits were obtained. Limited proteolysis of UPF2 (121–1031) led to the identification of three proteolytically stable fragments (one of ~30 kDa and two of ~45 kDa), each starting from residue 121 as determined by N-terminal sequencing (Supplementary Figure S1A). This information combined with sequence conservation and secondary structure [NPS@, (53)] and disorder [DisEMBL, (54)] predictions provided the basis for the design of different UPF2 constructs comprising a single MIF4G domain or two domains in tandem. Three well-behaved constructs were successfully crystallized: (i) residues 121–486 encompassing the MIF4G-1 domain, (ii) residues 455–757 encompassing the MIF4G-2 domain and (iii) the combined MIF4G-2/MIF4G-3 domains (residues 455–1054) (Supplementary Figure S1B). We solved the structure of UPF2 MIF4G-1 and MIF4G-2 domains at 2.6 and 2.4 Å resolution, respectively (Figure 1B and C), using SAD on seleno-methionine substituted protein crystals. Subsequently, the structure of the combined MIF4G-2/MIF4G-3 domains was determined at 5.4 Å resolution using molecular replacement. Crystallographic data are summarized in Supplementary Table S1.

UPF2 MIF4G-1 domain has an extended coiled coil and two capping helices

UPF2 MIF4G-1 contains the canonical 10-helix core MIF4G fold (residues 168–429) with five pairs of antiparallel α-helices forming an N-terminal four-helix bundle (helices h1–h4) and two parallel layers composed of three helices each (helices h5–h10) (Figure 1B and structure annotated sequence alignment in Supplementary Figure S2). In addition, the domain displays three interesting features: (i) helices h9 and h10 are highly elongated at their C- and N-terminus, respectively, to form a long coiled coil that protrudes away from the rest of the domain into the solvent, the loop at the extremity being poorly ordered but still possible to be modeled (Figure 1B, Supplementary Figure S3A); (ii) an additional helix (h8i) is inserted between helices h8 and h9 via two extended loops (residues 324–344) (Figure 1B, Supplementary Figure S3B); (iii) the core MIF4G fold is preceded at the N-terminus by an extended α-helix (hA, residues 121–149) followed by a loop and a second shorter α-helix (hB) connecting to helix h1 (residues 150–167) (Figure 1B); the N-terminus of helix hA is packed against helix h8i (Figure 1B and Supplementary Figure S3B and C).

MIF4G-1 helices hA, hB as well as their adjacent loops make extensive contacts with the 10-helix MIF4G core domain, in particular with helices h4, h6 and h8 (Figure 2A and B). This packing involves both hydrophobic and charged interactions. Notably, Arg137 and Arg141 on hA form hydrogen bonds and salt bridges with Asp317 and Glu271 on the core MIF4G domain, and Leu144 on hA interacts with Val223 and Val275 on h4 and h6, respectively (Figure 2A). Pro156, Phe160, Phe161 and Leu164 on hB form a hydrophobic patch interacting with Arg232 and Tyr233 on helix h4 (Figure 2B). Helix h8i and its adjacent loops pack against the MIF4G domain as well, establishing extensive hydrophobic and charged interactions with helices h6, h7, h8 and h9 (Supplementary Figure S3B). Helix h8i stabilizes the N-terminus of helix hA with mainly hydrophobic interactions (Supplementary Figure S3B). Overall, helices hA, hB and h8i create an extensive network of interactions with the 10 conserved helices of MIF4G-1 core MIF4G domain.

A highly conserved motif, 164-LDSSLKKNT-172 (Supplementary Figure S2), is located on the first turn of h1 and the loop immediately preceding it (Figure 2C and D). Mapping of phylogenetically conserved residues on the solvent-accessible surface of MIF4G-1 shows that the motif and a series of downstream conserved charged residues (Lys176, Lys177, Lys179, Asp192, Lys200, Glu204 and Glu211, Supplementary Figure S2) define a conserved patch at one extremity of the domain (Figure 2C). Asp165 and Lys170 side chains establish a salt bridge, while Ser167 N-terminally caps h1 by forming a hydrogen bond with the backbone amino group of Lys170, both contributing to fix the 3D configuration of the motif. Absolutely conserved Ser166, on the other

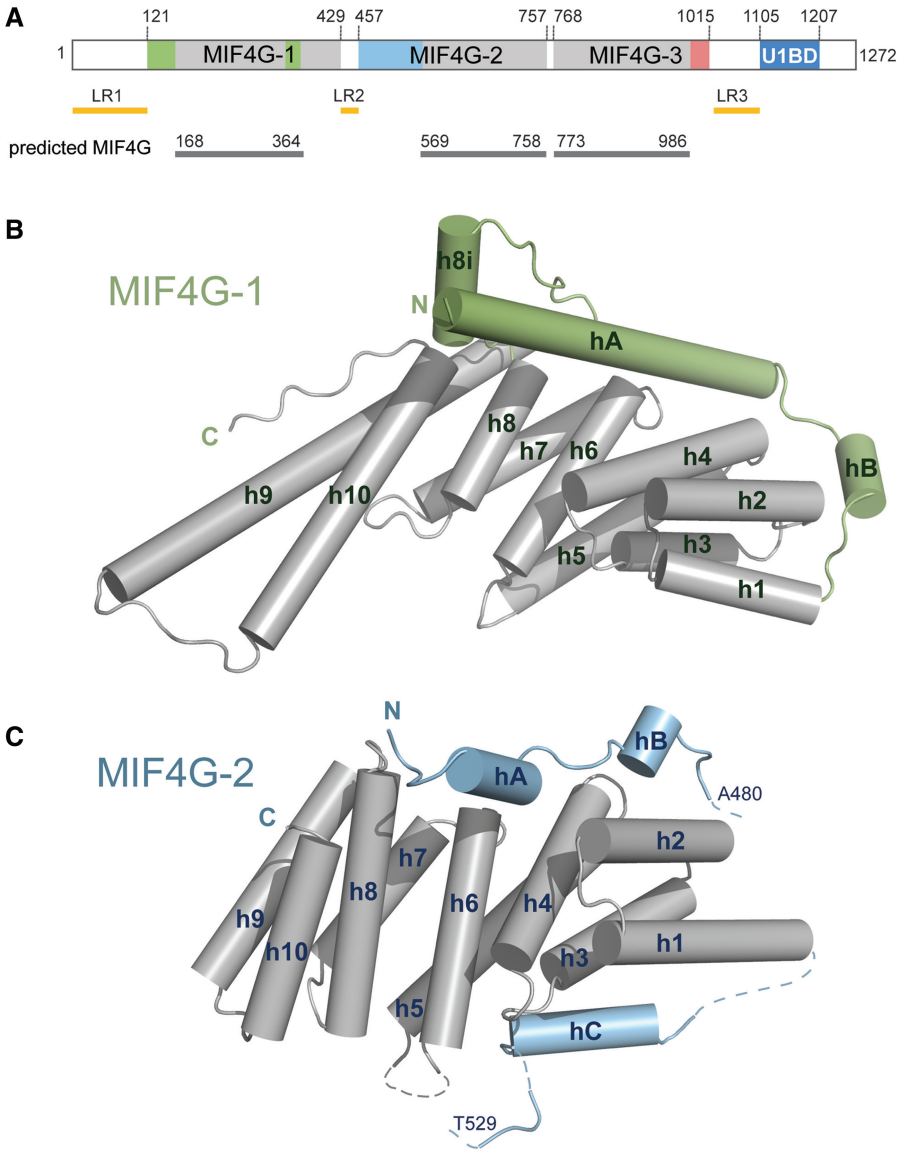


Figure 1. Crystal structures of UPF2 MIF4G-1 and MIF4G-2 domains. (**A**) Schematic representation of UPF2 indicating the domain boundaries (gray) of the conserved MIF4G domains according to SMART prediction. Colors (green, blue and pink) mark regions of the MIF4G domains that are not part of the classical 10-helix MIF4G domain fold. U1BD indicates the UPF1 binding domain. The N-terminal acidic low complexity region (residues 1–120) is indicated as LR1 (Linker Region 1). The disordered loop connecting MIF4G-1 and MIF4G-2 (residues 430–457) is indicated as LR2. The C-terminal acidic low complexity region (residues 1016–1104) connecting MIF4G-3 with U1BD is indicated as LR3 (**B**) Structure of MIF4G-1. The 10 helices (gray) of the classical MIF4G domain fold are numbered from 1 to 10. The N-terminal extension of the domain that folds into two helices (hA and hB) and the additional helix (h8i) inserted between h8 and h9 are depicted in green. The elongated helices h9 and h10 form a long coiled coil projecting from the otherwise compact domain. (**C**) Structure of MIF4G-2. The helix annotation is as in B with the N-terminal extension depicted in blue. It folds into three helices (hA, hB and hC), which are connected by partially disordered loops (shown as dashed lines). The long disordered loop between residues Ala480 and Thr529 is not represented.

hand, is solvent exposed and does not interact with other residues. Phosphorylation of one or both of the serines in this conserved motif has been reported to be required for NMD in yeast [(55), see discussion].

Residues 430–486 at the C-terminus of the MIF4G-1 crystallization construct are disordered in the crystal structure. This is consistent with the fact that the folded part of MIF4G-2 starts at residue 458 (see below), the region 430–457 being a flexible connecting loop between MIF4G-1 and MIF4G-2 (indicated as LR2 in Figure 1A) that is predicted to be disordered by DisEMBL and NPS@.

N-terminal capping helices also occur in the UPF2 MIF4G-2 domain

The crystal structure of UPF2 MIF4G-2 (residues 455–757) shows that the core MIF4G fold of 10 antiparallel helices (residues 561–756) (Figure 1C) is preceded by an ~100 amino acid region (residues 458–560) that contains additional elements essential for stability. The first part of this N-terminal region (458–478) folds into two short α-helices (hA and hB), which together with the connecting loops pack against helices h2, h4, h6 and

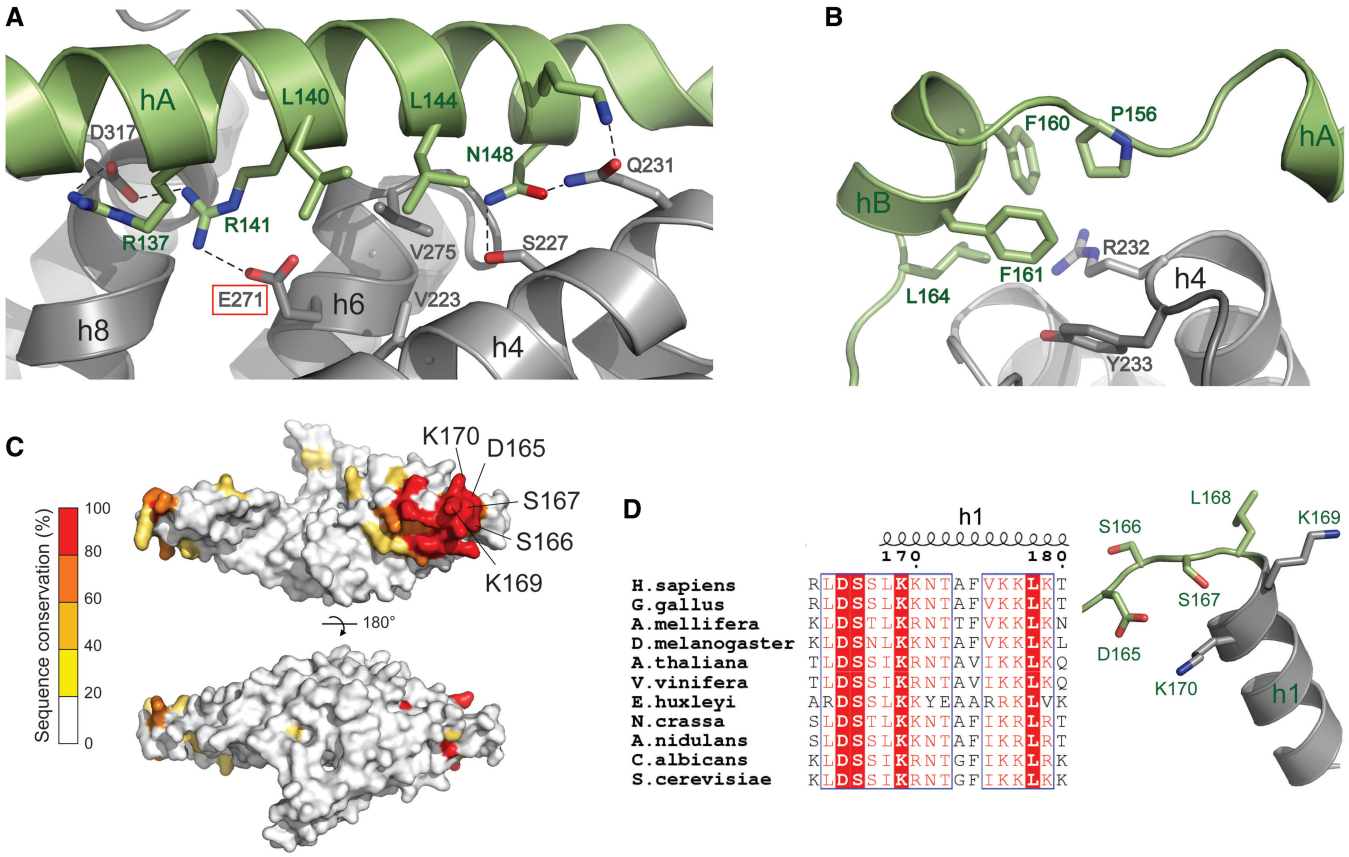


Figure 2. Interaction of the capping helices and adjacent loops with the MIF4G-1 core domain. Helix annotation and color code as in Figure 1B. (**A**) Helix hA interacts tightly with helices h4, h6 and h8 of the core domain. Amino acids involved in the interaction are shown in stick representation. The interaction involves hydrophobic interactions, hydrogen bonds and salt bridges. Notably, glutamate (E271) of the conserved FIGEL signature motif of MIF4G domains in helix h6 interacts with arginine 141 of helix hA. (**B**) Helix hB and its N-terminal loop interact via hydrophobic residues with the C-terminal tip of helix h4 of the conserved MIF4G fold. (**C**) Mapping of phylogenetically conserved residues on the solvent-accessible surface of MIF4G-1 based on the sequence alignment displayed in Supplementary Figure S2. Highly conserved residues (depicted in red) form a patch around the N-terminus of helix h1. For clarity a cartoon representation of MIF4G-1 in the same orientation is given in Supplementary Figure S3C. (**D**) Sequence alignment of the highly conserved 164-LDSSLKKNT-172 motif of UPF2 from representative species from yeast to human (left panel); 3D organization of residues 165–170 adjacent to helix h1 at its N-terminus (right panel).

h8 (Figure 3A and B). The interaction is mainly hydrophobic, involving Ile458, Trp459, Phe467 and Tyr468 of helix hA and Pro620, Phe621, Phe676 and Phe713 of h4, h6 and h8 (Figure 3A). Two salt bridges are also established between Glu460 and Arg712 and Arg716 on helix h8 (Figure 3A). Helix hB and its adjacent loops pack against helices h2 and h4 via hydrophobic interactions (Figure 3B). Residues 481–558 following helix hB are largely disordered in the crystal structure (apart from the short helix hC) and indeed this region is highly variable between species (Supplementary Figure S2). Similar to MIF4G-1, the N-terminal region stabilizes the interhelical packing of the MIF4G-2 domain helices. In fact, residues 458–478 are evolutionarily conserved within UPF2 (Supplementary Figure S2), and the expression of MIF4G-2/3 constructs lacking this capping region yields mostly insoluble protein (Supplementary Figure S3D). Notably, superposition of MIF4G-1 and MIF4G-2 shows that the capping helices hA and hB in the two domains pack against the same region of helices h2, h4, h6 and h8 (Supplementary Figure S4A).

Interaction of MIF4G-2 with MIF4G-3

We solved the structure of the combined MIF4G-2/MIF4G-3 domains by molecular replacement at 5.4 Å resolution, using the two separate domains as search models (Figure 4A). To validate this structure we also collected anomalous scattering data on seleno-methionine-substituted MIF4G-2/MIF4G-3 crystals. The position of the methionine residues in the molecular replacement two domain model coincided with the selenium atoms revealed in the anomalous difference map, confirming the placement of the MIF4G-2 and MIF4G-3 domains (Supplementary Figure S5). The structure of the combined MIF4G-2/MIF4G-3 domains reveals a rigid assembly with the two domains being orthogonally oriented with respect to each other (Figure 4A). The buried surface area is estimated to be ~750 and ~639 Å^2 for MIF4G-2 and MIF4G-3, respectively. The domains are connected via a short 11-amino acid-long linker (residues 757–767), which stretches the 23 Å distance between the C-terminus of MIF4G domain 2 and the N-terminus of MIF4G domain 3. The interaction

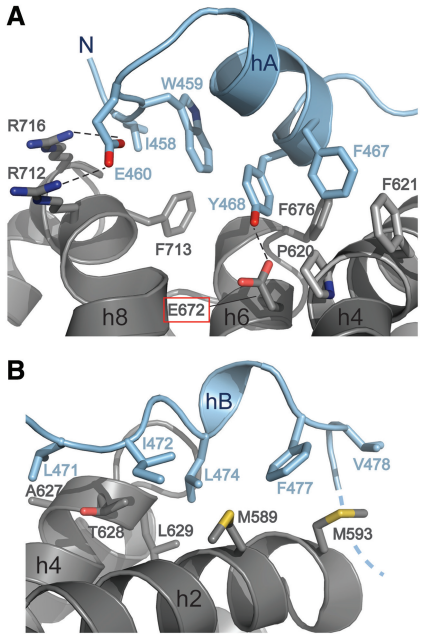
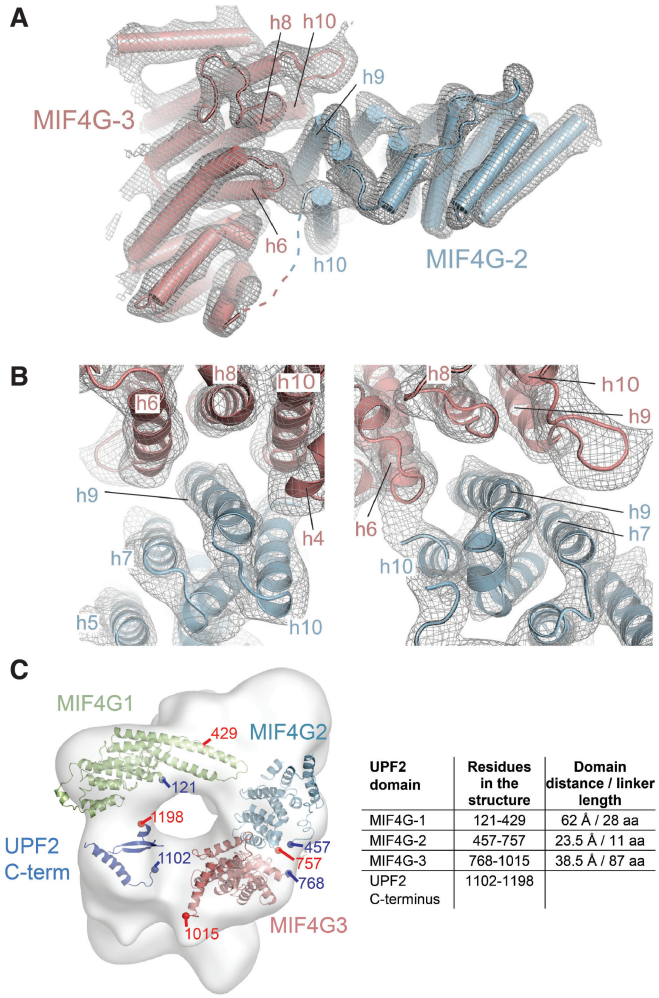


Figure 3. Interaction of the capping helices and adjacent loops with the MIF4G-2 core domain. Helix annotation and color code as in Figure 1C. (**A**) Helix hA and its N-terminal loop interact with helices h4, h6 and h8 of the core domain. Amino acids involved in the interaction are shown in stick representation. The interaction involves hydrophobic interactions, hydrogen bonds and salt bridges. Notably, glutamate (E672) of the conserved FIGEL signature motif in helix h6 interacts with tyrosine 468 of helix hA. (**B**) Helix hB and its adjacent loops interact via hydrophobic residues with helix h2 and h4 of the conserved MIF4G core.



UPF2 domain	Residues in the structure	Domain distance / linker length
MIF4G-1	121-429	62 Å / 28 aa
MIF4G-2	457-757	23.5 Å / 11 aa
MIF4G-3	768-1015	38.5 Å / 87 aa
UPF2 C-terminus	1102-1198	

Figure 4. Crystal structure of combined MIF4G-2 and MIF4G-3 domains of UPF2. (**A**) Overview showing the perpendicular orientation of the two interacting domains with MIF4G-2 in blue and MIF4G-3 in pale pink. The 2mFo-DF electron density obtained by molecular replacement (at 5.4 Å resolution and contoured at 1σ) within 3 Å from the MIF4G-2 and MIF4G-3 domains coordinates is depicted as a gray mesh. The dotted line represents the 10 residues linking h10 of MIF4G-2 to h1 of MIF4G-3. (**B**) Close-up of the interacting helices. Helix h9 of MIF4G-2 inserts into a cavity formed by helices h6, h8 and h10 of MIF4G-3. The N-terminal part of MIF4G-2 h9 interacts also with h4 of MIF4G-3 (left panel). Helix h10 of MIF4G-2 contacts h6 of MIF4G-3 as evidenced by the connecting electron density between the two helices (right panel). (**C**) Refined quasi-atomic model of human UPF2 using the UPF2 C-terminus [Clerici *et al.* (15)], UPF2 MIF4G-1 (this study) and MIF4G-2/MIF4G-3 (this study) crystal structures into the cryo-EM reconstruction of the UPF1/2/3-EJC complex [EMD-2048, Melero *et al.* (37)]. The table summarizes the UPF2 domain boundaries, the distances between the UPF2 domains in the quasi-atomic model as well as the length of the linker peptides that connect these domains with the subsequent domain.

is mediated mostly by MIF4G-2 helix h9, which inserts in a concave surface created by MIF4G-3 helices h6, h8 and h10 (Figure 4B).

The N-terminal part of MIF4G-2 helix h9 contacts the loop between MIF4G-3 helices h9 and h10 as well as the N-terminal part of helix h10. The C-terminal part of MIF4G-2 helix h9 interacts with the N-terminus of MIF4G-3 helix h4 (Figure 4B). An additional interaction is visible between the C-terminus of MIF4G-2 helix h10 and the N-terminus of MIF4G-3 h6.

The MIF4G-3 surface involved in the contact with MIF4G-2 is located opposite to the UPF3 binding site and therefore does not interfere with the UPF2-UPF3 interaction (Supplementary Figure S6). Superposition of MIF4G-3 in the UPF2/UPF3b complex (16) and in the MIF4G-2/MIF4G-3 structure indicates that the UPF3 RRM does not contact the UPF2 MIF4G-2 domain (Supplementary Figure S6).

UPF2 MIF4G domain arrangement in the context of the UPF-EJC complex

Recently, a cryo-EM structure of the UPF1/2/3-EJC complex has been determined and used to derive a quasi-atomic model of the entire complex using previously known crystal structures of UPF1 bound to UPF2-U1BD, UPF2 MIF4G-3 bound to UPF3b RRM and the RNA-bound EJC (37). The UPF2 MIF4G-1 and MIF4G-2 domains were modeled only as core MIF4G domains. The results suggest that the MIF4G domains of UPF2 adopt, together with the UPF1 zinc-knuckle (CH) domain, a ring-like structure, which forms the central scaffold for complex assembly (37). We placed the crystal structure of MIF4G-1 and the MIF4G-2/3 assembly into the EM density of the UPF-EJC complex (EMD-2048) using Chimera (50) (Figure 4C, Supplementary Figure S7). The updated quasi-atomic

model indicates that (i) the long coiled coil formed by helices h9 and h10 protruding from MIF4G-1 is in close proximity to and could contact the EJC, (ii) the loop between h9 and h10 contacts MIF4G-2 and (iii) the MIF4G-2 domain is positioned close to the Y14 subunit of the EJC likely forming a contact (Supplementary Figure S7). In conclusion, these observations support the proposed scaffolding function of UPF2 MIF4G domains (37) and indicate that UPF2 MIF4G-1 and MIF4G-2 domains are responsible for the positioning of the EJC in the UPF-EJC complex.

Role of MIF4G domains 1 and 2 on NMD *in vivo*

To evaluate the role of MIF4G-1 and MIF4G-2 in NMD we performed a UPF2 complementation assay in HeLa cells. In this assay the abundance of a NMD-competent triosephosphate isomerase (TPI) reporter mRNA harboring a PTC at position 48 is measured on siRNA silencing of endogenous UPF2 and rescue by transfection with different siRNA insensitive UPF2 expression constructs. Four UPF2 MIF4G-1 and/or MIF4G-2 deletion mutants were generated (Figure 5A): UPF2 ΔM1 and UPF2 ΔM2, lacking the core MIF4G-1 domain (residues 168–431) and the core MIF4G-2 domain (residues 569–758), respectively; UPF2 ΔN/M1 lacking MIF4G-1 and the preceding N-terminal region (residues 1–455) and UPF2 ΔN/M1M2 lacking MIF4G-1, MIF4G-2 and the preceding N-terminal region (residues 1–757). All four constructs are impaired in their ability to support NMD (Figure 5B lane 7–10 and Figure 5C) compared with wild-type UPF2 (Figure 5B lane 3 and Figure 5C). We noticed that these mutants show enhanced expression levels compared with wild-type UPF2 (Figure 5B), suggesting that their inability to support NMD is not the result of decreased protein stability. We also excluded that this effect is due to altered cellular localization since all the tested mutants localize in the cytoplasm as wild-type UPF2 (Supplementary Figure S8A). Next, we verified that the UPF2 mutants retained the ability to co-immunoprecipitate UPF1 and UPF3 (Figure 5D). Taking into account the differences in expression levels, UPF2 ΔM1 and ΔM2 show a virtually unaltered ability to interact with UPF1, whereas UPF2 ΔN/M1 and ΔN/M1M2 show enhanced co-immunoprecipitation of UPF1 (Figure 5D). UPF3 binding is unaltered in all four mutants. Altogether, these results demonstrate the essential role of MIF4G-1 and MIF4G-2 during NMD for reasons independent of protein stability, cellular localization or UPF3 and UPF1 binding. We also noticed that in the presence of these UPF2 constructs the reporter mRNA levels are higher than in the absence of transfected UPF2 (Figure 5B, lane 2 and Figure 5C). This dominant negative effect is likely due to the UPF2 mutants sequestering endogenous UPF1 and UPF3 and thus further reducing NMD in addition to endogenous UPF2 silencing.

We used the same UPF2 deletion constructs in a NMD tethering assay (24,51). In this assay, UPF2 is directly tethered to a beta-globin mRNA downstream of the PTC, thus bypassing the need for the interaction of UPF2 with UPF3-EJC for its recruitment to the mRNA. In this assay UPF2 ΔM1, ΔN/M1 and ΔN/M1M2 are able to support NMD with a partial impairment (Figure 5E lane 6–8 and Figure 5F) compared with wild-type UPF2 (Figure 5E lane 2 and Figure 5F). This indicates that when UPF2 is directly tethered to the mRNA, MIF4G-1 and MIF4G-2 are partially dispensable for its downstream interaction with the SURF complex and the triggering of UPF1 phosphorylation. Surprisingly, the UPF2 ΔM2 construct appears to be completely inactive in the tethering assay (Figure 5E lane 9 and Figure 5F). This is unexpected since the lack of the MIF4G-2 domain in the UPF2 ΔN/M1M2 construct did not severely affect NMD. Thus, the mere absence of MIF4G-2 cannot explain the failure of UPF2 ΔM2 to support NMD. Rather, the unusual proximity, in this construct, of the MIF4G-1 domain to the MIF4G-3 and UPF1 binding domains, may sterically interfere with UPF1 and SMG1 binding, thus affecting NMD.

The fitting of the UPF2 MIF4G domains in the EJC-UPF envelope (Supplementary Figure S7) suggests that the MIF4G-1 coiled coil contacts the EJC, thus possibly having a key role in the stabilization of the EJC-UPF complex. To test this possibility, we designed a UPF2 construct (Δcoiled coil) where the protruding region of MIF4G-1 coiled coil (residues 370–406) is replaced by a short linker (Ala-Gly-Gly-Gly-Ala) (Figure 5A). Both in the NMD complementation (Figure 5B lane 6 and Figure 5C) and tethering assay (Figure 5E, lane 5 and Figure 5F), the loss of MIF4G-1 coiled coil does not interfere with NMD, suggesting that this is not the only region involved in stabilizing the EJC-UPF complex.

SMG1 binds and phosphorylates UPF2 MIF4G-3

Previous work suggested that the UPF1 kinase SMG1 also interacts with UPF2 (19). We therefore tested whether various truncated UPF2 constructs are able to interact with full-length recombinant SMG1 by SPR. Wild-type SMG1 was immobilized via an N-terminal SBP-tag onto a streptavidin sensor chip, and the UPF2 constructs were injected. First, we determined the apparent dissociation constant $K_{D\text{-app}}$ using UPF2 containing all three MIF4G domains and the UPF1 binding domain (residues 121–1227). This measurement confirmed that SMG1 and UPF2 tightly interact with a $K_{D\text{-app}}$ of 37 ± 5 nM (Figure 6A). Subsequently, we asked which UPF2 domain interacts with SMG1. To this end, we tested the individual MIF4G domains as well as UPF2 constructs carrying different combinations of MIF4G domains. We were able to demonstrate that MIF4G domain 3 is sufficient for SMG1 interaction. In contrast, MIF4G-1 and MIF4G-2 do not show binding to SMG1 (Figure 6B). Moreover, we measured for the MIF4G-3/SMG1 complex a $K_{D\text{-app}}$ of 60 ± 5 nM (Supplementary Figure S9A), which constitutes a <2-fold reduction compared with the $K_{D\text{-app}}$ of 37 ± 5 nM obtained for UPF2 (121–1227) and SMG1. These results strongly suggest that MIF4G-3 is the main determinant for UPF2-SMG1 complex formation.

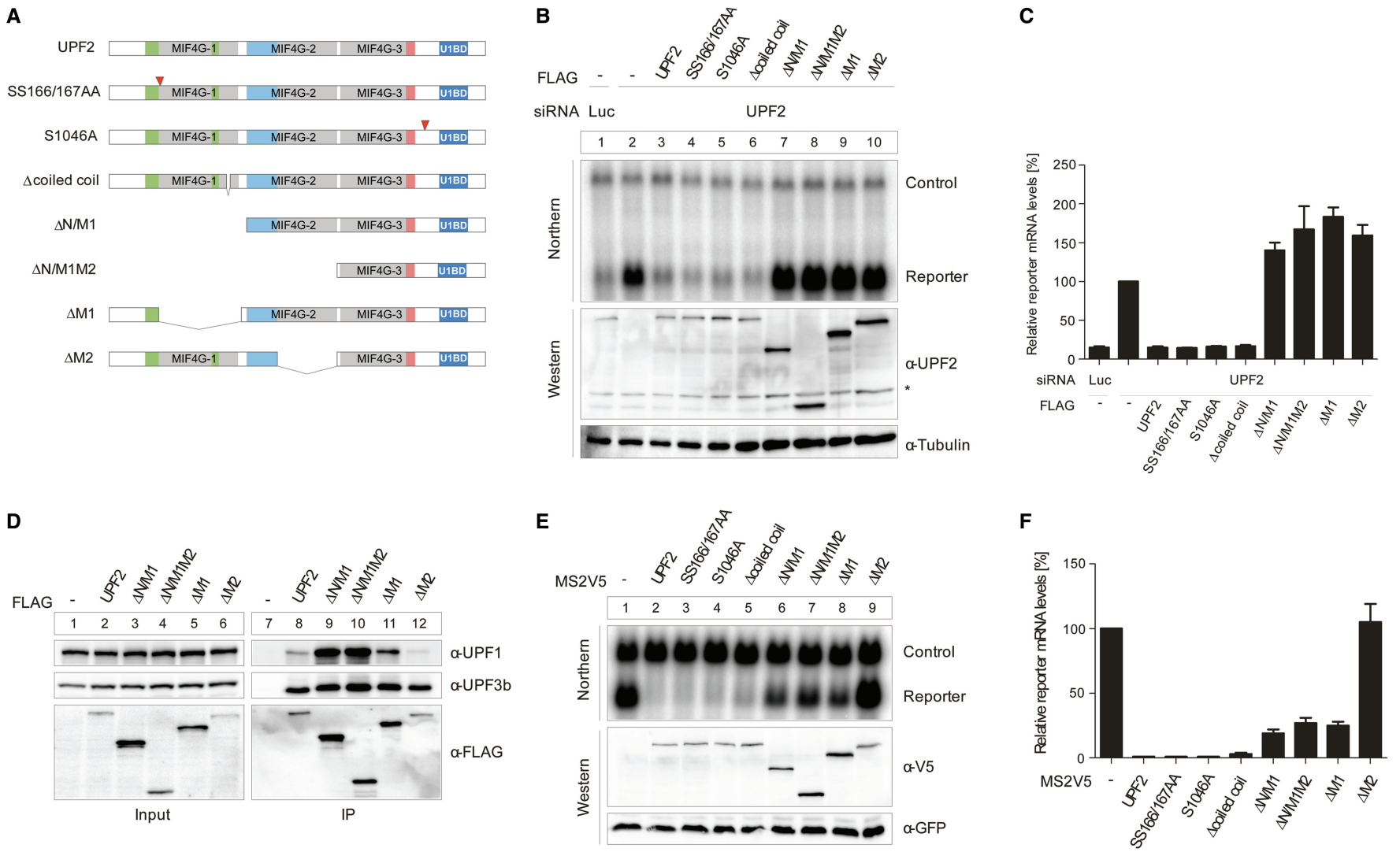


Figure 5. *In vivo* NMD tethering assays. UPF2 MIF4G-1 and -2 are required for NMD *in vivo*. (**A**) Schematic overview of UPF2 constructs used in the NMD analysis. UPF2 is represented as in Figure 1A; SS166/167AA and S1046A point mutant positions are indicated by red triangles. In the coiled coil deletion construct residues 370–406 were replaced by an Ala-Gly-Gly-Gly-Ala linker. (**B**) UPF2 complementation assay. The indicated siRNAs (Luc or UPF2) were transfected into HeLa cells. Next, plasmids encoding N-terminally FLAG-tagged UPF2 variants, the TPI-HBB PTC48 reporter mRNA and the LacZ-HBB control mRNA were transfected. Total RNA was analyzed by northern blotting. Protein samples were analyzed by western blotting using FLAG and tubulin (loading control) antibodies. The asterisk represents a nonspecific band. (**C**) Quantification of the relative reporter mRNA levels shown in the UPF2 complementation assay in (B), normalized to the control mRNA levels. (**D**) UPF1-UPF2-UPF3 co-immunoprecipitation assay. HeLa cells were transfected with the indicated FLAG-tagged expression plasmids. Immunoprecipitation was performed using FLAG beads and co-precipitated UPF1 and UPF3b were detected by western blotting using specific antibodies. (**E**) UPF2 tethering assay. HeLa cells were transfected with plasmids encoding the indicated N-terminally MS2V5-tagged UPF2 variants and co-transfected with plasmids for the reporter (β-globin 4MS2) and control (wt300+e3). Total RNA was isolated and analyzed by northern blotting. Expression levels of the tethered proteins were visualized by western blotting using a V5 antibody. GFP served as transfection control. (**F**) Quantification of the reporter mRNA levels shown in the UPF2 tethering assay in (E), normalized to the expression levels of the control mRNA.

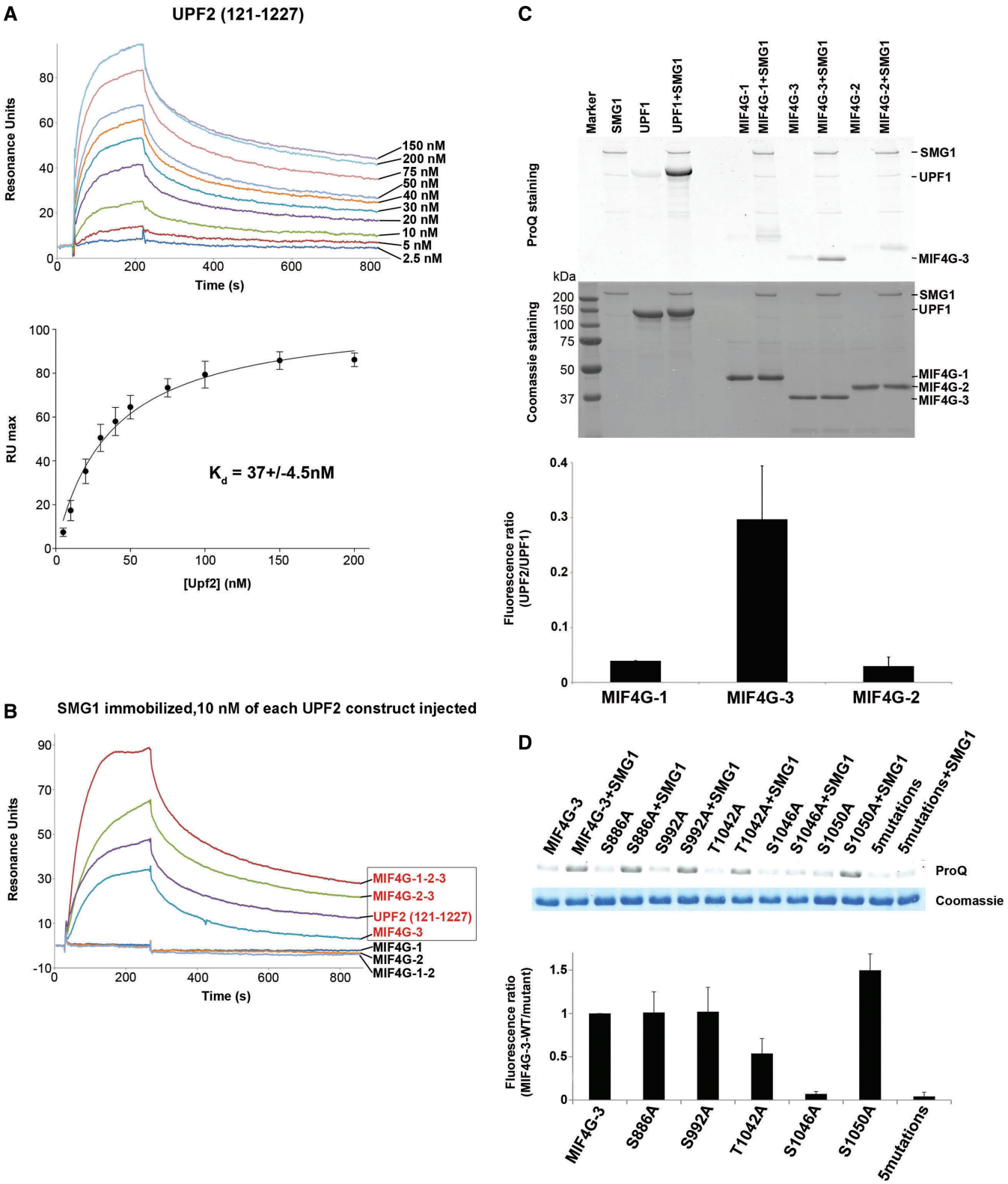


Figure 6. Characterization of the UPF2-SMG1 interaction. (**A**) Overlay of SPR sensograms of UPF2(121–1227) comprising all three MIF4G domains (above) interacting with immobilized SMG1. UPF2 (121–1227) was injected at concentrations ranging from 2.5 to 200 nM. The maximal resonance unit signal was plotted against the UPF2 concentration (below); the data points result from three independent experiments. The apparent dissociation constant $K_{D\text{-app}}$ was determined assuming a 1:1 interaction of SMG1 and UPF2. (**B**) SPR interaction profile of different UPF2 constructs with immobilized SMG1. All UPF2 constructs were injected at a concentration of 10 nM. UPF2 constructs containing the MIF4G-3 domain are labeled with red text color. Noninteracting constructs containing MIF4G-1 and/or MIF4G-2 are labeled with black text color. (**C**) *In vitro* phosphorylation experiments of UPF2 MIF4G constructs with SMG1 kinase. SMG1 was mixed with different UPF2 MIF4G domain constructs and with UPF1 as a positive control. The reaction mixture was analyzed by SDS-PAGE and Pro-Q Diamond

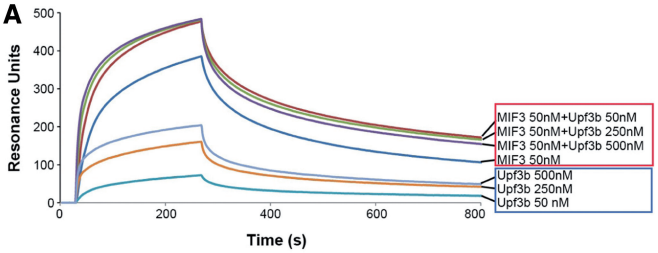
(continued)

Next, we investigated whether SMG1 is able to phosphorylate UPF2 MIF4G domains *in vitro*. We found that MIF4G-3, but not MIF4G-1 and MIF4G-2, is significantly phosphorylated by SMG1 in the presence of ATP (Figure 6C). MIF4G-3 shows about three times less phosphorylation as measured by Pro-Q staining compared with the known SMG1 substrate UPF1, which served as a positive control (Figure 6C). UPF1 has up to five Ser/Thr phosphorylation sites, which are recognized by SMG1 *in vivo* and *in vitro* (56). UPF2 (121–1227) carrying all three MIF4G domains is phosphorylated to a similar extent as MIF4G-3. We performed large-scale phosphorylation of UPF2 MIF4G-3, to purify the *in vitro* phosphorylated form. The pure domain was then subjected to mass spectrometry analysis. Three SMG1 phosphorylation sites were identified (Supplementary Table S2): Ser886, Ser992 and Thr1042, Ser1046 or Ser1050. Ser886 and Ser992 are located at the N-terminus of MIF4G-3 helices h7 and h11, respectively (Supplementary Figure S10). Residues Thr1042, Ser1046 and Ser1050 are located on the linker between the structured part of MIF4G-3 and the UPF2 C-terminus, which binds UPF1. In the UPF-EJC quasi-atomic model, Ser992, Thr1042, Ser1046 and Ser1050 are positioned at the free surface of the ring formed by UPF2 MIF4G domains (Supplementary Figure S10). Thus, they are likely to be exposed and accessible for SMG1 phosphorylation in the UPF-EJC complex. Next, each individual residue has been substituted to alanine and tested for *in vitro* phosphorylation by SMG1. All alanine mutants except Ser1046 are phosphorylated to a similar extent as the wild-type MIF4G-3 domain (Figure 6D). Moreover, mutation of all five Ser/Thr led to a similar phosphorylation signal as the single S1046A mutant (Figure 6D), indicating that Ser1046 is the main phosphosite for SMG1 in the UPF2 MIF4G-3 construct (residues 761–1054), which we used as substrate in our assays.

Next, we performed NMD complementation and tethering assays in HeLa cells to assess whether the UPF2 S1046A mutation would affect NMD *in vivo*. In both experiments, the S1046A mutation does not interfere with NMD *in vivo* (Figure 5B lane 5 and Figure 5E lane 4). To rule out the possibility that a single phosphorylation site mutation is not sufficient to produce a detectable effect on NMD efficiency, we tested a UPF2 construct carrying alanine mutations for all five identified phosphorylation sites and tested it in a complementation assay. Also for this construct we did not detect any effect on NMD *in vivo* (Supplementary Figure S8B). In addition, the serine residues 166 and 167, which are located in the highly conserved motif of MIF4G-1 and have been reported to be important for NMD in yeast (55), were both mutated to alanines. The mutation of these residues did not impair UPF2 in complementation as well as in tethering assays (Figure 5B lane 4 and Figure 5E lane 3).

SMG1 and UPF3b can simultaneously bind to UPF2 MIF4G-3 domain

Since UPF3b and SMG1 both bind to UPF2 MIF4G-3 domain, we tested whether this binding was mutually exclusive using two different approaches. First, we performed a competition experiment where the binding of MIF4G-3 at constant concentration (50 nM) to SMG1 was measured by SPR in the presence of UPF3b. Increasing concentrations of UPF3b (50, 250 and 500 nM) do not decrease the binding of UPF2/UPF3b to SMG1 (Figure 7A). This indicates that UPF3b and SMG1 do not compete for UPF2 binding and that the UPF2/UPF3b-SMG1 complex can form. In the case of overlapping binding sites, i.e. competition, addition of UPF3b to UPF2 would reduce the SPR signal. Rather, formation of the UPF3-MIF4G-3 complex leads to an increased signal in SPR (Figure 7A). Second, we



B

Protein constructs used for SPR experiments	KD with SMG1 and standard deviation
UPF2 (120-1227)	37 ± 5
UPF2 MIF4G-3	60 ± 5
UPF2 MIF4G-3 + UPF3b RRM (42-143)	34 ± 2
UPF3b RRM (42-143)	331 ± 31

Figure 7. UPF2 MIF4G-3 binding to UPF3 and SMG1 is not mutually exclusive. (**A**) SPR interaction profile of UPF2 with SMG1 in the presence of increasing concentration of UPF3b RRM domain (red box), and profile of UPF3b RRM domain interacting with SMG1 (blue box). UPF2 was injected at a concentration of 50 nM in the presence of 0, 50, 250, 500 nM UPF3b. UPF3b alone was injected a concentrations ranging from 50 to 500 nM. (**B**) Apparent dissociation constants ($K_{D\text{-app}}$) of the UPF protein constructs and SMG1 determined by SPR experiments, which identify the MIF4G-3 domain as the SMG1 binding domain and the existence of a ternary MIF4G-3/UPF3b/SMG1 complex.

Figure 6. Continued
Staining (Life technologies) (top) staining only phosphorylated proteins. SMG1 is auto-phosphorylated and therefore stains in the experiment as well. The Coomassie-stained gel (middle) shows all proteins present the experiments. The phosphorylation signal after Pro-Q staining was quantified with a Typhoon scanner (bottom). Each bar presents the average of three independent experiments. (**D**) As for (**C**), SMG1 was mixed with different UPF2 MIF4G-3 mutants to assess *in vitro* phosphorylation. The reaction mixture was analyzed by SDS-PAGE and Pro-Q Diamond (top) and Coomassie staining (middle). The phosphorylation level of each mutant was quantified using the ratio of its fluorescence signal to the wild-type fluorescence signal (bottom). Each bar represents the average of four independent experiments.

measured by SPR the apparent dissociation constant between SMG1 and the preformed MIF4G-3/UPF3b complex. The apparent dissociation constant of SMG1 and UPF2 MIF4G-3 in complex with UPF3b ($K_{D\text{-app}} = 34 \pm 2$ nM) and of SMG1 and MIF4G-3 alone ($K_{D\text{-app}} = 60 \pm 5$ nM) is similar (Figure 7B and Supplementary Figure S11), confirming that the UPF3b RRM domain does not compete or interfere with the UPF2-SMG1 complex formation.

As a control, we tested the binding of UPF3b RRM domain (residues 42–143) to SMG1. Compared with UPF2, UPF3b shows an ~10-fold weaker interaction with SMG1 ($K_{D\text{-app}} = 331 \pm 31$ nM) (Figure 7B and Supplementary Figure S9B) with fast on and off rates. The UPF3b domain is highly charged at neutral pH (UPF3b isoelectric point is 8.8), and thus the observed binding could be due to nonspecific charged interactions.

DISCUSSION

UPF2 MIF4G domain 1 and 2 folds are stabilized by N-terminal capping helices

UPF2 is composed of three conserved tandem MIF4G domains and a C-terminal UPF1 interacting domain (Figure 1A). Structural information is available for the MIF4G-3 interacting with the RRM domain of UPF3b (16) and of the C-terminus of UPF2 in complex with UPF1 (15). Here, we present the crystal structures of the remaining UPF2 folded domains, the N-terminal MIF4G-1 and MIF4G-2 domains. In both cases, we found that the canonical 10-helical MIF4G fold is augmented by an N-terminal extension that packs onto the core and is essential for overall domain stability (Figure 1). The additional helices in the N-terminus, in particular helix hA, are located in a similar position in the MIF4G-1 and -2 domains and pack against helices h2, h4, h6 and h8 of the MIF4G core involving both hydrophobic and charged interactions (Figures 2 and 3). Consistent with the multiple interactions of the capping helices with the MIF4G core, deletion of the N-terminal extension renders the domains aggregation-prone, indicating that it has an important stabilizing function. We note that while UPF2 MIF4G-3 does not have an N-terminal extension (although its contact with MIF4G-2 may play a stabilizing role), it does have an extra C-terminal helix (h11) added to the core MIF4G fold. We provide a more extensive discussion of the structural conservation of MIF4G domains in the Supplementary Material.

MIF4G-1 contains unique elongated helices

In addition to the N-terminal extension, MIF4G-1 contains two elongated helices (h9 and h10) (Figure 1). This is a striking and phylogenetically conserved feature of the UPF2 MIF4G-1 domain, although a similar extended coiled coil also occurs in the third, C-terminal, MIF4G-like domain of CBP80 but formed by helices equivalent to h8 and h9 and thus pointing in a different direction (40). When fitting the MIF4G-1 domain into the EM density of the UPF-EJC complex (37), we observe that the loop connecting h9 and h10 is in close proximity and likely to contact MIF4G-2 (Supplementary Figure S7). Thus, this protrusion may be important for stabilizing the correct 3D arrangement of the UPF2 MIF4G domains in the UPF-EJC complex, MIF4G-1 and MIF4G-2 otherwise being connected by a flexible linker. Moreover, MIF4G-1 h9 and h10 are in a position to establish contacts to the EJC in the complex (Supplementary Figure S7). Therefore, MIF4G-1 could contribute to the postulated scaffolding function of UPF2 (37) by positioning the EJC in the DECID complex. We tested this hypothesis with *in vivo* NMD assays (Figure 5) but found that deletion of the extended part of h9 and h10 does not affect NMD.

MIF4G domain arrangement in UPF2

The crystal structure of the tandem MIF4G-2/MIF4G-3 domains shows that they form a rigid assembly with a substantial total buried surface area of ~1400 Å^2. The two domains are oriented perpendicular to each other with helices h9 and h10 of MIF4G-2 interacting with helices h4, h6, h8 and h10 of MIF4G-3 (Figure 4). Interestingly, in MIF4G-1 and 2, these helices (h4, h6, h8 and partly h10) are contacted by the N-terminal extension (hA, hB and adjacent loops). In contrast to the observed rigid juxtaposition of MIF4G-2 and MIF4G-3, limited proteolysis analysis of UPF2 (Supplementary Figure S1A) suggests that MIF4G-1 is flexibly linked with respect to MIF4G-2/MIF4G-3 in free UPF2 and only becomes fixed in position within the UPF-EJC complex.

Sequence analysis shows that proteins usually contain two or more MIF4G or MIF4G-like (HEAT or ARM) domains in tandem connected by predicted unstructured linkers (57,58). This is the case for CBP80, eIF4G and UPF2. The structure of CBP80 provides an example of three combined MIF4G/MIF4G-like domains (40) where the middle HEAT domain is the central core of the protein and the N-terminal MIF4G and the C-terminal HEAT domains pack against it, forming a compact assembly. CBP80 thus shows a globular shape different from the ring-like arrangement of UPF2 MIF4G domains, which is suggested by the UPF-EJC EM reconstruction (37). Consistently, the superposition of UPF2 MIF4G-2/MIF4G-3 to CBP80 MIF4G domains does not indicate any obvious similarity in the relative domain organization of the two proteins. Therefore, even though proteins containing tandem MIF4G/MIF4G-like domains may have a common ancestor [as suggested for eIF4G and CBP80 (58)], the structural arrangement of the MIF4G domains can clearly differ from protein to protein.

UPF2 MIF4G-1 and -2 scaffold the UPF-EJC complex but have no direct function in interaction with the SURF complex

The EM structure of the UPF-EJC complex shows that UPF2 forms a ring-like scaffold for the assembly of the UPF-EJC complex (37). Our refined quasi-atomic model indicates that MIF4G-1 and 2 are key bridging factors between the EJC and the UPF1 CH domain and the

UPF2 MIF4G-3 domain, respectively (Supplementary Figure S7).

By UPF2 complementation assays we tested the impact of UPF2 MIF4G-1 and MIF4G-2 deletion on NMD efficiency. All tested deletions completely impaired the ability of UPF2 to activate NMD of a reporter mRNA (Figure 5B and C), confirming the absolute requirement of MIF4G-1 and MIF4G-2 for NMD in cells. The same constructs were also used in a UPF2 tethering assays. In this situation it is likely that only the ability of the UPF2 variants to interact with the SURF complex and to activate UPF1 phosphorylation, which ultimately triggers mRNA degradation, is tested. Contrary to the complementation assays, UPF2 ΔM1, ΔN/M1 and ΔN/M1M2 were only partially impaired to support NMD in comparison with the wild type (Figure 5E and F). Thus, by the complementary use of these two assays we were able to uncouple MIF4G-1 and 2 functions in scaffolding the UPF-EJC complex (complementation) and in activating the downstream events (tethering). The diverging outcome of the two assays strongly indicates that MIF4G-1 and MIF4G-2 are strictly necessary for NMD to promote the correct assembly of the EJC and the UPF proteins, but are partially dispensable for the interaction of UPF2 with the SURF complex and triggering of UPF1 phosphorylation.

MIF4G domain 3 plays a central role in the DECID complex

The critical 'point-of-no-return' in NMD is the activation of SMG1 kinase to phosphorylate UPF1 (31,59). This is achieved through association of the SURF complex with the UPF2/3-EJC complex to form the decay inducing complex (DECID) (19). Details of the UPF1 interaction with UPF2 have been revealed by radiographic crystallography (15) and with the UPF2/3-EJC complex by cryo-EM (37). Immunoprecipitation experiments indicate a direct interaction between UPF2 and the C-terminal region of SMG1 that contains the kinase domain (19). Here, we identify UPF2 MIF4G-3 as the interaction site with SMG1 kinase (Figure 6). We measured by SPR an apparent dissociation constant of 37 nM for UPF2 (121–1227) and SMG1 (Figure 6A), which is only slightly increased to 60 nM when the MIF4G-3 domain alone binds SMG1 (Supplementary Figures S9A and Figure 7B). No interaction between SMG1 and MIF4G domains 1 and 2 could be detected *in vitro* (Figure 6B). Our finding is further supported by the *in vivo* tethering assays that indicate that the deletion of the N-terminal region of UPF2 does not strongly interfere with UPF2 interaction with the SURF complex (Figure 5A).

The UPF2 MIF4G-3 domain is a relatively small domain that interacts stably with both MIF4G-2 (Figure 4) and the UPF3b RRM domain (16). Moreover, according to the UPF-EJC cryo-EM structure, UPF2 MIF4G-3 is also in close proximity with the UPF1 CH domain (37). Therefore, we investigated whether UPF2 MIF4G-3/UPF3b complex formation would interfere with UPF2 MIF4G-3/SMG1 interaction. The measured apparent dissociation constant of MIF4G-3 and SMG1 in the presence of UPF3b RRM domain (34 nM) is comparable with that of MIF4G-3 alone (60 nM), suggesting that UPF3b does not compete with UPF2 for SMG1 binding and that the trimeric UPF2/UPF3b/SMG1 complex can exist also *in vivo*.

SMG1 kinase is suggested to be downregulated by the SMG8/SMG9 heterodimer (12). SMG9 has been shown to bind to the N-terminal HEAT-repeat region of SMG1 and may act as a recruitment platform for SMG8, which, upon binding, induces conformational changes in SMG1 and inhibits its kinase activity (33). UPF2 has been reported to interact with the C-terminal region of SMG1 carrying the kinase domain (19), and therefore UPF2 likely does not compete with SMG8/SMG9 for binding the N-terminus of SMG1. Consequently, UPF2 could act directly on the kinase domain to activate SMG1.

UPF2 is a SMG1 kinase substrate

Here we provide further evidence for a direct interaction of UPF2 with the SMG1 kinase domain. We observed that UPF2 MIF4G-3 is phosphorylated by SMG1 kinase *in vitro* (Figure 6C). A combination of mass spectrometry and *in vitro* mutational analyses identified three phosphosites: Ser886, S992 and Ser1046 but only the latter one is quantitatively phosphorylated *in vitro* (Figure 6D). Ser1046 is not part of a canonical SQ site. However, a Gln is present before the Ser1046 and this residue is located in a Glu-rich region. These two elements have been found in ATM/ATR phosphosites identified by mass spectrometry (60). According to the quasi-atomic model derived from the UPF-EJC complex reconstruction [Melero *et al.* (37)], Ser1046 is positioned on a flexible loop and thus likely to be exposed. Furthermore, it is positioned close to the UPF1 phosphosites and thus could be accessible by SMG1 (Supplementary Figure S10). Therefore, the identification of an *in vitro* phosphorylation site in MIF4G-3 confirms the direct interaction of UPF2 MIF4G-3 with the SMG1 kinase domain, possibly leading to activation of SMG1 to phosphorylate UPF1. However, mutation of Ser1046 or of all five identified phosphorylation sites are not sufficient to interfere with NMD in cells, neither in complementation assays nor in tethering assays (Figure 5). Consistent with our findings, human UPF2 has been reported to be a phosphoprotein (36), but it has not been identified as a SMG1 kinase substrate *in vivo*, nor has any functional role for human UPF2 phosphorylation been proposed. In contrast, UPF2 binding to SMG1 kinase is essential to trigger NMD (19).

In contrast, a study in *Saccharomyces cerevisiae* showed that yeast Upf2p is phosphorylated *in vivo* (55), most likely including serines 32 and/or 33 in the highly conserved motif near the beginning of the MIF4G-1 domain (30-LDSSIKRNT in yeast, see above and Figure 2C and D). Mutational analyses showed that the residues 21-DSS are important for interaction of Upf2p with the yeast-specific NMD factor Hrp1p and for triggering NMD *in vivo* (55). If this motif adopts the same configuration as observed in the human MIF4G-1 structure, only Ser32 (equivalent to Ser166 in human UPF2) is solvent exposed and could be phosphorylated. In fact,

Ser32 is phylogenetically absolutely conserved. However, an interaction partner for this motif in higher eukaryote UPF2 remains to be identified. UPF2 constructs with mutations, S166A, S166D (data not shown) and SS166/167AA (Figure 5) in this motif were tested in tethering and complementation experiments. However, no effect on NMD was observed (Figure 5B and E), suggesting that the intact motif is not required for DECID complex formation. Therefore, it remains to be shown whether UPF2 phosphorylation by SMG1 (which has no homologue in yeast) or by any other kinase has a significant role in NMD in higher eukaryotes.

In summary, we elucidate here the structure of UPF2 MIF4G domains 1 and 2 and show that they exert a role in the structural organization of the EJC-UPF complex. UPF2 MIF4G domain 3 interacts with the SMG1 kinase domain and UPF3b and plays the key role in DECID complex formation and SMG1 activation, which triggers NMD.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

We thank the ESRF for providing access to synchrotron beamlines and Drs Thibaut Crépin and Andrew McCarthy for their help with crystallographic data collection. The technical platforms of the Partnership for Structural Biology (PSB) were extensively used, notably the robotic crystallization facility. We thank Prof. Elena Conti for sharing the quasi-atomic model of the UPF-EJC complex and Jens Lykke-Andersen for providing anti-UPF1 and anti-UPF2 antibodies. A.D. acknowledges access to the Mammalian Expression Platform at University of Oxford through the FP7 I3 project P-CUBE (www.p-cube.eu) and practical support by Drs Yuguang Zhao and Radu Ariecescu. Dr Franck Fieschi is acknowledged for access to the Biacore 3000 and the EMBL Proteomics core facility for analysis of SMG1 phosphorylated peptides. M.C. carried out all protein crystallization experiments and collected radiographic data; M.C. and S.C. solved and refined the crystal structures. A.D. performed the biophysical characterization of the SMG1/UPF2/UPF3b interactions. V.B. and N.H.G. designed and performed *in vivo* NMD assays. M.C., C.S. and S.C. wrote the manuscript.

FUNDING

European Molecular Biology Laboratory EIPOD postdoctoral programme [fellowship to A.D.]; European Research Council [ERC Starting Grant 281331 ComplexNMD to C.S.]; Swiss National Science Foundation [Sinergia grant CRSII3_136254, Structure/Function analyses of PIKKs to C.S.]; International Graduate School in Development Health and Disease IGSDHD [fellowship to V.B.]; the Deutsche Forschungsgemeinschaft [SFB635, project B6 to N.H.G.]. Funding for open access charge: EMBL.

Conflict of interest statement. None declared.

REFERENCES

1. Behm-Ansmant,I., Kashima,I., Rehwinkel,J., Sauliere,J., Wittkopp,N. and Izaurralde,E. (2007) mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. *FEBS Lett*., **581**, 2845–2853.
2. Doma,M.K. and Parker,R. (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*, **440**, 561–564.
3. Isken,O. and Maquat,L.E. (2007) Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev*., **21**, 1833–1856.
4. Rebbapragada,I. and Lykke-Andersen,J. (2009) Execution of nonsense-mediated mRNA decay: what defines a substrate? *Curr. Opin. Cell. Biol*., **21**, 394–402.
5. Shyu,A.B., Wilkinson,M.F. and van Hoof,A. (2008) Messenger RNA regulation: to translate or to degrade. *EMBO J*., **27**, 471–481.
6. Nicholson,P. and Muhlemann,O. (2010) Cutting the nonsense: the degradation of PTC-containing mRNAs. *Biochem. Soc. Trans*., **38**, 1615–1620.
7. Bhuvanagiri,M., Schlitter,A.M., Hentze,M.W. and Kulozik,A.E. (2010) NMD: RNA biology meets human genetic medicine. *Biochem. J*., **430**, 365–377.
8. Anders,K.R., Grimson,A. and Anderson,P. (2003) SMG-5, required for C.elegans nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. *EMBO J*., **22**, 641–650.
9. Rehwinkel,J., Letunic,I., Raes,J., Bork,P. and Izaurralde,E. (2005) Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA*, **11**, 1530–1544.
10. Behm-Ansmant,I. and Izaurralde,E. (2006) Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. *Genes Dev*., **20**, 391–398.
11. Isken,O. and Maquat,L.E. (2008) The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet*., **9**, 699–712.
12. Yamashita,A., Izumi,N., Kashima,I., Ohnishi,T., Saari,B., Katsuhata,Y., Muramatsu,R., Morita,T., Iwamatsu,A., Hachiya,T. *et al.* (2009) SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. *Genes Dev*., **23**, 1091–1105.
13. Izumi,N., Yamashita,A., Iwamatsu,A., Kurata,R., Nakamura,H., Saari,B., Hirano,H., Anderson,P. and Ohno,S. (2010) AAA+ proteins RUVBL1 and RUVBL2 coordinate PIKK activity and function in nonsense-mediated mRNA decay. *Sci. Signal*., **3**, ra27.
14. Chen,Y.H., Su,L.H. and Sun,C.H. (2008) Incomplete nonsense-mediated mRNA decay in Giardia lamblia. *Int. J. Parasitol*., **38**, 1305–1307.
15. Clerici,M., Mourao,A., Gutsche,I., Gehring,N.H., Hentze,M.W., Kulozik,A., Kadlec,J., Sattler,M. and Cusack,S. (2009) Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J*., **28**, 2293–2306.
16. Kadlec,J., Izaurralde,E. and Cusack,S. (2004) The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat. Struct. Mol. Biol*., **11**, 330–337.
17. Czaplinski,K., Ruiz-Echevarria,M.J., Paushkin,S.V., Han,X., Weng,Y., Perlick,H.A., Dietz,H.C., Ter-Avanesyan,M.D. and Peltz,S.W. (1998) The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev*., **12**, 1665–1677.
18. Ivanov,P.V., Gehring,N.H., Kunz,J.B., Hentze,M.W. and Kulozik,A.E. (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J*., **27**, 736–747.
19. Kashima,I., Yamashita,A., Izumi,N., Kataoka,N., Morishita,R., Hoshino,S., Ohno,M., Dreyfuss,G. and Ohno,S. (2006) Binding of

a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev.*, **20**, 355–367.
20. Ponting,C.P. (2000) Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. *Trends Biochem. Sci.*, **25**, 423–426.
21. Chakrabarti,S., Jayachandran,U., Bonneau,F., Fiorini,F., Basquin,C., Domcke,S., Le Hir,H. and Conti,E. (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell*, **41**, 693–703.
22. Chamieh,H., Ballut,L., Bonneau,F. and Le Hir,H. (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.*, **15**, 85–93.
23. Serin,G., Gersappe,A., Black,J.D., Aronoff,R. and Maquat,L.E. (2001) Identification and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein and Upf3 protein (Caenorhabditis elegans SMG-4). *Mol. Cell. Biol.*, **21**, 209–223.
24. Gehring,N.H., Neu-Yilik,G., Schell,T., Hentze,M.W. and Kulozik,A.E. (2003) Y14 and hUpf3b form an NMD-activating complex. *Mol. Cell*, **11**, 939–949.
25. Bono,F., Ebert,J., Lorentzen,E. and Conti,E. (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell*, **126**, 713–725.
26. Buchwald,G., Ebert,J., Basquin,C., Sauliere,J., Jayachandran,U., Bono,F., Le Hir,H. and Conti,E. (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc. Natl Acad. Sci. USA*, **107**, 10050–10055.
27. Lejeune,F. and Maquat,L.E. (2005) Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.*, **17**, 309–315.
28. Le Hir,H., Gatfield,D., Izaurralde,E. and Moore,M.J. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.*, **20**, 4987–4997.
29. Kim,V.N., Kataoka,N. and Dreyfuss,G. (2001) Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science*, **293**, 1832–1836.
30. Lykke-Andersen,J., Shu,M.D. and Steitz,J.A. (2000) Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell*, **103**, 1121–1131.
31. Yamashita,A., Ohnishi,T., Kashima,I., Taya,Y. and Ohno,S. (2001) Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.*, **15**, 2215–2228.
32. Lempiainen,H. and Halazonetis,T.D. (2009) Emerging common themes in regulation of PIKKs and PI3Ks. *EMBO J.*, **28**, 3067–3073.
33. Arias-Palomo,E., Yamashita,A., Fernandez,I.S., Nunez-Ramirez,R., Bamba,Y., Izumi,N., Ohno,S. and Llorca,O. (2012) The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. *Genes Dev.*, **25**, 153–164.
34. Fukuhara,N., Ebert,J., Unterholzner,L., Lindner,D., Izaurralde,E. and Conti,E. (2005) SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol. Cell*, **17**, 537–547.
35. Huntzinger,E., Kashima,I., Fauser,M., Sauliere,J. and Izaurralde,E. (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA*, **14**, 2609–2617.
36. Chiu,S.Y., Serin,G., Ohara,O. and Maquat,L.E. (2003) Characterization of human Smg5/7a: a protein with similarities to Caenorhabditis elegans SMG5 and SMG7 that functions in the dephosphorylation of Upf1. *RNA*, **9**, 77–87.
37. Melero,R., Buchwald,G., Castano,R., Raabe,M., Gil,D., Lazaro,M., Urlaub,H., Conti,E. and Llorca,O. (2012) The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3′ end. *Nat. Struct. Mol. Biol.*, **19**, 498–505.
38. He,F., Brown,A.H. and Jacobson,A. (1997) Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol. Cell. Biol.*, **17**, 1580–1594.
39. Craig,A.W., Haghighat,A., Yu,A.T. and Sonenberg,N. (1998) Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. *Nature*, **392**, 520–523.
40. Mazza,C., Ohno,M., Segref,A., Mattaj,I.W. and Cusack,S. (2001) Crystal structure of the human nuclear cap binding complex. *Mol. Cell*, **8**, 383–396.
41. Pestova,T.V., Shatsky,I.N. and Hellen,C.U. (1996) Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol. Cell. Biol.*, **16**, 6870–6878.
42. Schutz,P., Bumann,M., Oberholzer,A.E., Bieniossek,C., Trachsel,H., Altmann,M. and Baumann,U. (2008) Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **105**, 9564–9569.
43. Kabsch,W. (2010) Xds. *Acta Crystallogr.*, **66**, 125–132.
44. McCoy,A.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
45. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr.*, **66**, 486–501.
46. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **53**, 240–255.
47. Vagin,A.A., Steiner,R.A., Lebedev,A.A., Potterton,L., McNicholas,S., Long,F. and Murshudov,G.N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr.*, **60**, 2184–2195.
48. Sheldrick,G.M. (2008) A short history of SHELX. *Acta Crystallogr. A*, **64**, 112–122.
49. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.*, **364**, 215–230.
50. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
51. Gehring,N.H., Kunz,J.B., Neu-Yilik,G., Breit,S., Viegas,M.H., Hentze,M.W. and Kulozik,A.E. (2005) Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol. Cell*, **20**, 65–75.
52. Steckelberg,A.L., Boehm,V., Gromadzka,A.M. and Gehring,N.H. (2012) CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep.*, **2**, 454–461.
53. Combet,C., Blanchet,C., Geourjon,C. and Deleage,G. (2000) NPS@: network protein sequence analysis. *Trends Biochem. Sci.*, **25**, 147–150.
54. Linding,R., Jensen,L.J., Diella,F., Bork,P., Gibson,T.J. and Russell,R.B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure*, **11**, 1453–1459.
55. Wang,W., Cajigas,I.J., Peltz,S.W., Wilkinson,M.F. and Gonzalez,C.I. (2006) Role for Upf2p phosphorylation in Saccharomyces cerevisiae nonsense-mediated mRNA decay. *Mol. Cell. Biol.*, **26**, 3390–3400.
56. Okada-Katsuhata,Y., Yamashita,A., Kutsuzawa,K., Izumi,N., Hirahara,F. and Ohno,S. (2011) N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res.*, **40**, 1251–1266.
57. Aravind,L. and Koonin,E.V. (2000) Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. *Genome Res.*, **10**, 1172–1184.
58. Marintchev,A. and Wagner,G. (2005) eIF4G and CBP80 share a common origin and similar domain organization: implications for the structure and function of eIF4G. *Biochemistry*, **44**, 12265–12272.
59. Grimson,A., O'Connor,S., Newman,C.L. and Anderson,P. (2004) SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. *Mol. Cell. Biol.*, **24**, 7483–7490.
60. Stokes,M.P., Rush,J., Macneill,J., Ren,J.M., Sprott,K., Nardone,J., Yang,V., Beausoleil,S.A., Gygi,S.P., Livingstone,M. *et al.* (2007) Profiling of UV-induced ATM/ATR signaling pathways. *Proc. Natl Acad. Sci. USA*, **104**, 19855–19860.

Structural and functional analysis of the three MIF4G domains of nonsense mediated decay factor UPF2

Supplementary Material

Marcello Clerici[1,2,4], Aurélien Deniaud[1,2,4], Volker Boehm[3], Niels H. Gehring[3], Christiane Schaffitzel[1,2,5] and Stephen Cusack[1,2,5]

[1] European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

[2] Unit of Virus Host-Cell Interactions, Univ. Grenoble Alpes-EMBL-CNRS, UMI 3265, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

[3] University of Cologne, Institute for Genetics, Zuelpicher Str. 47a, 50674 Cologne, Germany

[4] These authors contributed equally.

[5] Corresponding authors.

Contents:

Supplementary text

Structural conservation of MIF4G domains.

A global pair-wise structure comparison (DaliLite, http://www.ebi.ac.uk/Tools/dalilite/) of UPF2 MIF4G-1 and MIF4G-2 domains with other MIF4G domain structures of eIF4G, CBP80, UPF2 MIF4G-3 and MIF4Gdb (PDB accession codes 1HU3, 1H6K, 1UW4 and 2I2O respectively) (Bae, 2010; Kadlec, 2004; Marcotrigiano, 2001; Mazza, 2001) indicates that core MIF4G domains share a similar overall fold with RMSDs between 2.0 and 4.0 Å, despite low sequence identity of between 10 and 23% (Supplementary Table 3, Supplementary Figure S4B). UPF2 MIF4G-1 has the highest structural similarity to eIF4G MIF4G domain (Cα RMSD 2.9 Å, residues 752-986, 13% identity), whereas MIF4G-2 is most similar to MIF4G-3 (Cα RMSD 2.5 Å, residues 768-983, 16% identity) (Supplementary Table 3). The identity in primary sequence does not correlate with the similarity in the structure, in agreement with the observation that structure-based alignments do not show any strictly conserved residues in MIF4G domains with the exception of the FIGEL motif (see below). Rather, they contain a pattern of hydrophobic residues important for interhelical packing (Kadlec, 2004). The structural comparison also shows a conservation of the position of the helices, with the exception of the N-terminal four-helix bundle of MIF4Gdb. The helix length is also generally conserved, with the exception of UPF2 MIF4G-1 helices h9 and h10 which are elongated (Supplementary Figure S4B). In contrast, the loops connecting the helices vary considerably. UPF2 MIF4G-1 and CBP80 contain an additional helix inserted between helices h8 and h9 and helices h9 and h10 respectively. All MIF4G domains share a conserved ‘FIGEL’ sequence motif on helix h6, with the hydrophobic residues being involved in interhelical packing and the glutamic residue pointing to the exterior (Kadlec, 2004; Letunic, 2002). In MIF4G-1, the glycine residue is mutated into an alanine (268-FIAEL-272), while the canonical motif is present in MIF4G-2 (669-FIGEL-673). In UPF2

MIF4G-1 and MIF4G-2, the exposed glutamic residue of the FIGEL motif is involved in the binding of the capping helix hA: in MIF4G-1, Glu271 establishes a salt bridge with Arg141 of helix hA (Figure 2A); in MIF4G-2, Glu672 interacts with helix hA forming a hydrogen bond with the hydroxyl group of a conserved tyrosine, Tyr468 (Figure 3A). In the light of these observations, we re-examined the context of the first, canonical, MIF4G domain of human CBP80. The ten helices of this domain are followed by a long proline-rich linker (residues 245-308), containing three short helices, which completely encircles the domain (Figure 1 of (Mazza, 2001) and Supplementary Figure S12A) making many specific contacts with it (Supplementary Figure S12B). Some regions of this linker make analogous interactions to that observed in the UPF2 MIF4G-2 structure. In particular Asp133 of the 'FIGEL' motif (129-FLSDL-134 in human CPB80) makes a hydrogen bond with Tyr253 of the linker (Supplementary Figure S12B). Thus, it seems possible to distinguish two categories of MIF4G domains. On one hand those such as the MIF4G domains of eIF4G, MIF4Gdb and UPF2 MIF4G-3 which are stable as 'naked' canonical MIF4G domains. On the other hand, those where the interhelical packing is stabilised by peptide extensions to the core domain, perhaps to limit flexibility of the helical bundle, such as the first two MIF4G domains of UPF2 and that of CBP80. Very recently, the structure of the MIF4G domain of Not1 has been published as part of the yeast Ccr4-Not complex (Basquin, 2012; Fabian, 2013; Petit, 2012). This shows another example of stabilisation of the core helical fold by N- and C-terminal extensions that wrap around the helices.

Materials and Methods

Protein expression, purification and crystallisation

DNA fragments encoding his-tagged UPF2(121-1031), UPF2(121-486), UPF2(455-757) and UPF2(455-1054) were cloned into the pProExHTb expression vector (Invitrogen). Plasmids were transformed into *E. coli* BL21Star(DE3), and the cells grown overnight at 20°C after induction with 1 mM IPTG. The proteins were purified by immobilised Ni^{2+} affinity chromatography. After His-tag removal using TEV protease (leaving a Gly-Ala-Met-Gly extension at the N-terminus), the UPF2 constructs were loaded a second time onto a Ni^{2+} column. The last purification step included size exclusion chromatography (Superdex 200) in 20 mM Tris pH 7, 150 mM NaCl and 2 mM DTT followed by concentration of the proteins. Crystallisation trials were performed with a Cartesian robot which makes 100nl+100nl drops, and positive hits conditions were refined by the hanging-drop vapour diffusion technique. UPF2-MIF4G-1 (121-486) crystals were obtained in 100 mM bicine pH 9, 100 mM NaCl, 11% PEG 6000 and at a protein concentration of 16 mg/ml; UPF2 MIF4G-2 (455-757) crystals were obtained in 100 mM MES pH 6.0, 19% PEG 3350 and at a protein concentration of 8 mg/ml; UPF2 MIF4G-2/MIF4G-3 (455-1054) crystals were obtained in 100 mM HEPES pH 7.0, 100 mM NH_4SO_4, 200 mM $NaKHPO_4$ pH 7, 840 mM sodium malonate, 1% w/v PEG MME 2K and at a protein concentration of 10 mg/ml. Expression of seleno-methionine labelled MIF4G-1, MIF4G-2 and MIF4G-2/MIF4G-3 was carried out in *E. coli* BL21Star(DE3) growing in M9 minimal medium. Thirty minutes before induction (1 mM IPTG) the cells were supplemented with a cocktail of aminoacids containing seleno-methionine (final concentration 60 mg/l). The labelled proteins were purified and crystallised in the same conditions as the native ones.

Crystallographic data collection and structure determination

Crystallographic statistics are given in Supplementary Table I. All data collection was performed at 100K at the European Synchrotron Radiation Facility with crystals being briefly soaked in a solution containing mother liquor and 20% glycerol as cryoprotectant and snap-frozen into liquid nitrogen. The data were integrated with XDS (Kabsch, 2010) and analysed with CCP4i. Molecular replacement was performed with PHASER (McCoy, 2007), model building with COOT (Emsley, 2010) and refinement with REFMAC5 (Murshudov, 1997; Vagin, 2004). As molecular replacement with previously known MIF4G domains did not work, the structures of both MIF4G-1 and MIF4G-2 were solved *de novo* by selenomethionine (SeMet) SAD, using SHELXD (Sheldrick, 2008) to find sites and SHARP (De la Fortelle, 1997) for refinement and phasing. For the MIF4G-1 structure, the SeMet data was the best quality and therefore used for refinement. For the low resolution combined MIF4G-2/MIF4G-3 domain structure, molecular replacement using PHASER (McCoy, 2007) using the individual domains, gave a unambiguous unique solution with log likelihood gain (LLG) of 220. This was confirmed by correspondence of predicted methionine positions with anomalous difference peaks using data from MIF4G-2/MIF4G-3 crystals grown with SeMet. Due to the low resolution, no further refinement was performed.

Generation of the quasi-atomic UPF-EJC model

Starting from the EM reconstruction of the UPF-EJC complex (EMD-2048) and the corresponding quasi-atomic model (Melero, 2012), we replaced the UPF2 MIF4G domains and the UPF3b RRM domain by our crystal structure of MIF4G-1 domain and by the MIG4G-2/3/UPF3b RRM complex. We used Chimera (Pettersen, 2004) to obtain the best correlation coefficient for the placement of the domains into the density. To avoid a clash with MIF4G-1 and MIF4G-2 (both are larger than the MIF4G homology model used in the original atomic model), the EJC had to be moved by 14 Å away from UPF2. Similarly, the

UPF1 CH-domain was moved by 5 Å away from UPF2 MIF4G-3 (Supplementary Figure S7). The figures were generated in PyMOL (DeLano Scientific). In the rendering of the cryo-EM map, the density cutoff was set for the display of the envelope to represent ~130% of the a priori estimated volume.

In vivo NMD assays

The β-globin 4MS2 plasmid construct and the transfection control (wt300+e3) for the tethering assay were described previously (Gehring, 2005; Gehring, 2003). HeLa cells were grown in DMEM and transfected by calcium phosphate precipitation in 6-well plates with 0.8 µg of MS2-UPF2 fusion constructs, 0.5 µg of the control plasmid, 2 µg of the 4MS2 reporter plasmid, and 0.5 µg of a GFP expression plasmid. Total RNA was extracted with TRI Reagent (Sigma) and analyzed by northern blotting as described (Gehring, 2005; Gehring, 2003). Signals were quantified using a Typhoon Trio (GE Healthcare) and percentages ± standard deviations were calculated from three independent experiments.

For NMD complementation experiments HeLa cells were grown in 6 cm plates and transiently transfected with 100 pmol siRNA (UPF2 target sequence: 5'-CACGTTGTGGATGGAGTGTTA-3'; Luc target sequence: 5'-CGTACGCGGAATACTTCGATT-3') using Lipofectamine RNAiMAX (Life Technologies). On the next day, cells were split 1:2 into 10 cm plates and 24 h later transfected with 200 pmol siRNA. The next day, cells were seeded into 6-well plates and transfected by calcium phosphate precipitation with 0.5 µg of a GFP expression plasmid, 0.8 µg of expression plasmid for FLAG or siRNA insensitive (targeting sequence 5'-CACGTTGTGGATGGAGTGTTA-3' replaced by 5'-CATGTGGTTGACGGCGTCCTG-3') UPF2, SS166/167AA, S1046A and Δcoiled coil, 0.4 µg plasmid for ΔN/M1 and ΔN/M1M2,

0.6 µg for ΔM1 and ΔM2, 1.5 µg control plasmid (LacZ-HBB) and 2 µg plasmid encoding the reporter mRNA (TPI-HBB) (Steckelberg, 2012), harbouring a stop codon at position 48.

For UPF2 localisation HeLa cells were transfected with 600 ng of pCI-mVenus-UPF2 plasmids. Cells were plated on coverslips 24 hours post transfection. 48 hours after transfection cells were fixed with 3.7% formaldehyde. Images were recorded on a FV1000 confocal microscope (Olympus).

Immunoprecipitation, immunoblotting and antibodies

SDS-polyacrylamide gel electrophoresis and immunoblot analysis was performed using protein samples derived either from TRI Reagent extractions (Figure 5D) or from lysates (RIPA buffer supplemented with protease inhibitor) of parallel transfections (Figure 5B). For coimmunoprecipitation UPF2 mutants were expressed as FLAG-tagged proteins in HeLa cells. Fusion proteins were immunoprecipitated with FLAG-beads (Sigma-Aldrich) and co-purified proteins analysed by immunoblotting. The antibodies against tubulin (T6074) and FLAG (F7425) were from Sigma, the antibody against V5 (18870) was from QED Bioscience, the antibody against GFP (ab290) was from Abcam and the antibodies against UPF1 and UPF2 were kindly provided by Jens Lykke-Andersen. UPF3B antiserum was raised in rabbits by Eurogentech against a C-terminal fragment of UPF3B (300-483) and affinity purified.

SMG1 purification

SMG1 was expressed and purified as described in (Izumi, 2010) with some modifications. Lysed cells in SMG1 buffer (20 mM Hepes pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, 0.5 mM DTT, 0.05% tween-20, 5% glycerol) supplemented with protease and phosphatase inhibitors (Roche) were ultracentrifuged at 100 000x g for 30 minutes. The

supernatant was mixed with streptavidin beads and incubated for 2 hours at 4 °C with mixing. After washing with SMG1 buffer containing 100 mM NaCl and then 500 mM NaCl, SMG1 was eluted with SMG1 buffer with 1.5 mM biotin. Subsequently, SMG1 was concentrated and applied on a Superose-6 column equilibrated with SMG1 buffer. The fractions corresponding to monomeric SMG1 were pooled, concentrated, snap-frozen and stored at -80 °C.

SMG1 kinase assay

0.2 µg of SMG1 were mixed with 1 µg of either UPF1-FL, UPF2 (121-486), UPF2 (455-757) or UPF2 (761-1054) in SMG1 buffer containing 2 mM DTT, 5 mM $MnCl_2$ and 5 mM ATP. The mixture was incubated 1 h at room temperature. Proteins were separated on 12 % SDS-PAGE gels first stained with Pro-Q Diamond Staining (Life Technologies) and then with Coomassie staining. The Pro-Q stained gel was revealed with a Typhoon scanner using 532 nm and 580 nm as excitation and emission wavelength, respectively.

Surface Plasmon Resonance

SPR experiments were performed on a BIAcore 3000 using SA sensor chips (GE-Healthcare). One flow-cell was not functionalized to be used as a background control. The second flow-cell was functionalized with purified SMG1 to a density of about 2000 RU. The running buffer was 20 mM Hepes pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, 0.5 mM DTT, 0.05% Tween-20. UPF2/UPF3b solutions were injected at 25 µl/min during 3 or 4 min followed by a 10 min dissociation phase. The surfaces were regenerated by 1 minute injection of 0.5 M and 1 M NaCl solutions. The data were analyzed with the Biacore evaluation software by subtracting both the control flow cell and the buffer injection curve. The apparent equilibrium dissociation constants ($K_{D\text{-}app}$) were determined using RU values 10 seconds before the end of

the association phase for all curves (RU_{max}). These RU_{max} were plotted as a function of protein concentration and fitted assuming a one binding site model.

Phosphosites identification by mass spectrometry

500 µg of UPF2 (761-1054) in SMG1 buffer containing 2 mM DTT, 5 mM $MnCl_2$ and 5 mM ATP were mixed with or without 10 µg of SMG1. The mixtures were incubated 5 h at room temperature. UPF2 and SMG1 were then separated using a Superose-6 column equilibrated with SMG1 buffer containing 500 mM NaCl. The peak corresponding to UPF2 was collected and concentrated to 100 µl. Samples were then diluted with 50mM NH_4HCO_3 to obtain 20 µL of 0.5 µg/µl of protein , which was reduced with DTT (50 mM, 2 µL) for 30 min at 56 °C and alkylated with iodacetamide (110 mM, 2µL) for 20 min at room temperature in the dark. Incubation with trypsin (0.1 µg enzyme/10 µg protein) was carried out overnight at 37 °C. Digestion was stopped by adding 4 µl of 10% trifluoracetic acid. Prior to analysis by liquid chromatography - mass spectrometry (LC-MS/MS), the peptides were diluted to 0.1 µg/µl with 0.1% formic acid. Peptides were separated using the nanoAcquity UPLC system (Waters) fitted with a trapping (nanoAcquity Symmetry C18, 5µm, 180 µm x 20 mm) and an analytical column (nanoAcquity BEH C18, 1.7µm, 75µm x 200mm). The outlet of the analytical column was coupled directly to an LTQ Orbitrap Velos (Thermo Fisher Scientific) using the Proxeon nanospray source. The mass spectrometric raw data was processed using MaxQuant (version 1.1.1.36) (Cox, 2008).

Supplementary references

Bae, E., E. Bitto, C.A. Bingman, J.G. McCoy, G.E. Wesenberg and G.N. Phillips, Jr. (2010). Crystal structure of an eIF4G-like protein from Danio rerio. *Proteins* **78**(7): 1803-6.

Basquin, J., V.V. Roudko, M. Rode, C. Basquin, B. Seraphin and E. Conti (2012). Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. *Mol Cell* **48**(2): 207-18.

Cox, J. and M. Mann (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**(12): 1367-72.

De la Fortelle, E., J.J. Irwin and G. Bricogne (1997). SHARP: A Maximum-Likelihood Heavy-Atom Parameter Refinement and Phasing Program for the MIR and MAD Methods. *Crystallographic computing*.

Emsley, P., B. Lohkamp, W.G. Scott and K. Cowtan (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4): 486-501.

Fabian, M.R., F. Frank, C. Rouya, N. Siddiqui, W.S. Lai, A. Karetnikov, P.J. Blackshear, B. Nagar and N. Sonenberg (2013). Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. *Nat Struct Mol Biol* **20**(6): 735-9.

Gehring, N.H., J.B. Kunz, G. Neu-Yilik, S. Breit, M.H. Viegas, M.W. Hentze and A.E. Kulozik (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol Cell* **20**(1): 65-75.

Gehring, N.H., G. Neu-Yilik, T. Schell, M.W. Hentze and A.E. Kulozik (2003). Y14 and hUpf3b form an NMD-activating complex. *Mol Cell* **11**(4): 939-49.

Izumi, N., A. Yamashita, A. Iwamatsu, R. Kurata, H. Nakamura, B. Saari, H. Hirano, P. Anderson and S. Ohno (2010). AAA+ proteins RUVBL1 and RUVBL2 coordinate PIKK activity and function in nonsense-mediated mRNA decay. *Sci Signal* **3**(116): ra27.

Kabsch, W. (2010). Xds. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 125-32.

Kadlec, J., E. Izaurralde and S. Cusack (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol* **11**(4): 330-7.

Letunic, I., L. Goodstadt, N.J. Dickens, T. Doerks, J. Schultz, R. Mott, F. Ciccarelli, R.R. Copley, C.P. Ponting and P. Bork (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* **30**(1): 242-4.

Marcotrigiano, J., I.B. Lomakin, N. Sonenberg, T.V. Pestova, C.U. Hellen and S.K. Burley (2001). A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol Cell* **7**(1): 193-203.

Mazza, C., M. Ohno, A. Segref, I.W. Mattaj and S. Cusack (2001). Crystal structure of the human nuclear cap binding complex. *Mol Cell* **8**(2): 383-96.

McCoy, A.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**: 658-674.

Melero, R., G. Buchwald, R. Castano, M. Raabe, D. Gil, M. Lazaro, H. Urlaub, E. Conti and O. Llorca (2012). The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3' end. *Nat Struct Mol Biol* **19**(5): 498-505, S1-2.

Murshudov, G.N., A.A. Vagin and E.J. Dodson (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**(Pt 3): 240-55.

Petit, A.P., L. Wohlbold, P. Bawankar, E. Huntzinger, S. Schmidt, E. Izaurralde and O. Weichenrieder (2012). The structural basis for the interaction between the CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. *Nucleic Acids Res* **40**(21): 11058-72.

Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng and T.E. Ferrin (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**(13): 1605-12.

Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr A* **64**(Pt 1): 112-22.

Steckelberg, A.L., V. Boehm, A.M. Gromadzka and N.H. Gehring (2012). CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep* **2**(3): 454-61.

Vagin, A.A., R.A. Steiner, A.A. Lebedev, L. Potterton, S. McNicholas, F. Long and G.N. Murshudov (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2184-95.

	UPF2 MIF4G-1 (121-486) SeMet data	**UPF2 MIF4G-2 (455-757) Native data**	**UPF2 MIF4G-2 (455-757) SeMet data**	**UPF2 MIF4G-2/3 (455-1054) Native data**	**UPF2 MIF4G-2/3 (455-1054) SeMet data**
Data collection (date)	07/07/2007	15/02/2009	07/03/2009	10/04/2008	07/10/2008
Beamline (ESRF)	BM14	ID23-2	ID14-4	ID14-4	ID14-4
Wavelength (Å)	0.9788	0.8726	0.9795	0.9395	0.9795
Space group	$P6_5$	$P2_12_12_1$	$P2_12_12_1$	$I2_12_12_1$	$I2_12_12_1$
Cell dimensions *a*, *b*, *c* (Å)	101.1, 101.1, 158.9	40.9, 56.6, 142.2	41.3, 56.7, 145.8	92.5, 132.1, 216.6	95.4, 132.4, 218.5
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å) *	50-2.60 (2.70-2.60)	50-2.35 (2.40-2.35)	50-3.10 (3.10-3.20)	50-5.4 (5.8-5.4)	50-6.0 (6.0-6.3)
R_{merge}*	0.089 (0.63)	0.066 (0.81)	0.104 (0.848)	0.068 (0.68)	0.058 (0.466)
$I / \sigma I$*	10.9 (1.7)	13.8 (1.9)	10.5 (1.7)	10.7 (2.0)	9.2 (1.7)
Completeness (%)*	99.1 (99.3)	99.7 (99.4)	99.9 (100)	95.9 (97.0)	97.4 (98.0)
Redundancy*	2.86 (2.37)	3.95 (4.09)	3.98 (4.00)	3.58 (3.65)	1.94 (1.95)
Refinement					
Resolution (Å) *	50-2.6 (2.67-2.60)	50-2.35 (2.41-2.35)			
No. reflections work/free	26788/1428	13638/723			
R_{work}*	0.205 (0.327)	0.206 (0.286)			
R_{free}*	0.251 (0.420)	0.251 (0.348)			
No. atoms	5115	1990			
Protein atoms	5061 (2 copies)	1958 (1 copy)			
Water molecules	54	32			
Average B-factors (Å^2)	55.2	53.4			
R.m.s. deviations					
Bond lengths (Å)	0.007	0.012			
Bond angles (°)	1.059	1.462			
Ramachandran Plot**					
Favoured (%)	96.1	94.8			
Allowed (%)	99.2	100			

* Values in parentheses are for highest-resolution shell. ** Molprobity http://molprobity.biochem.duke.edu

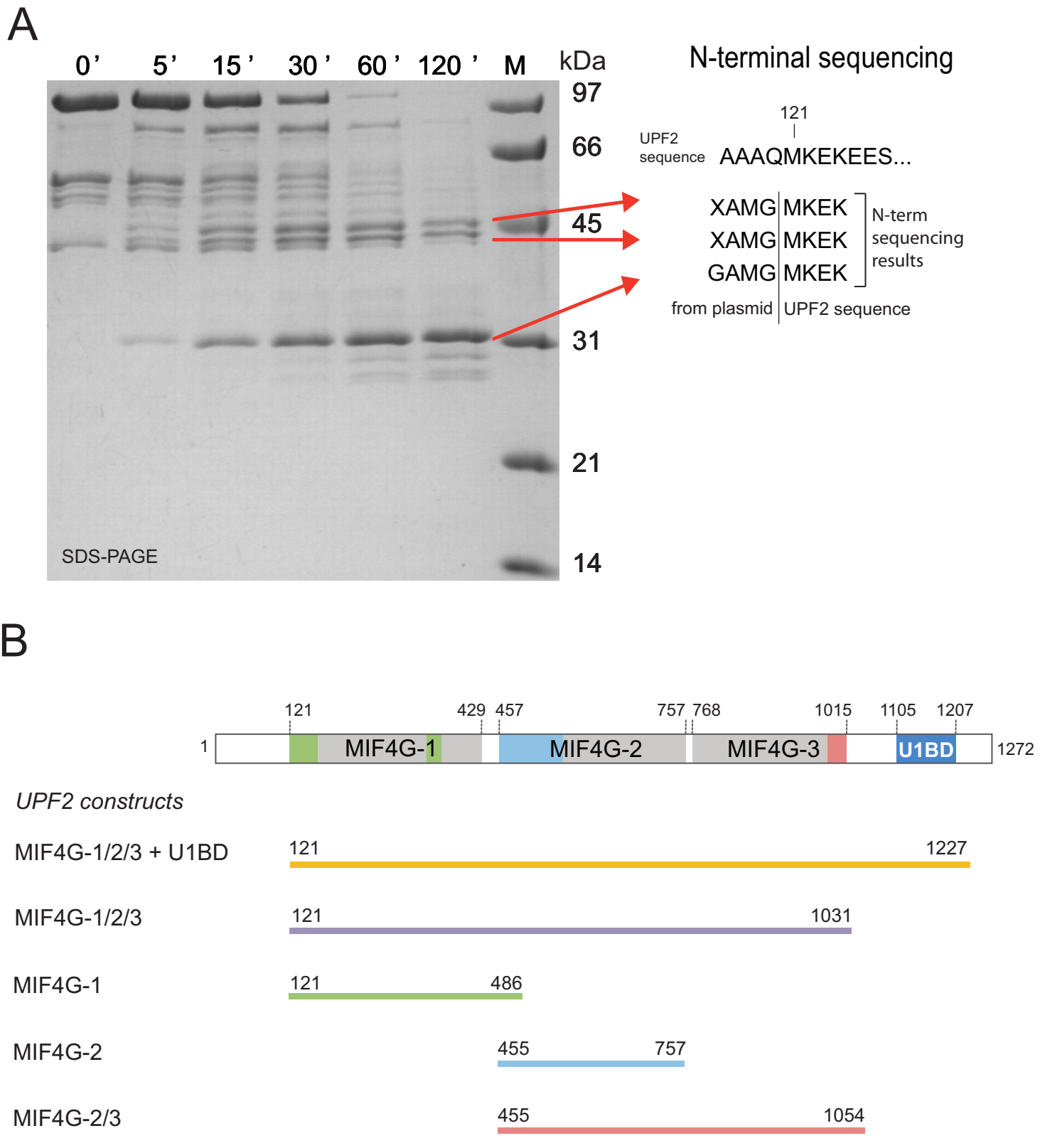
Supplementary Table 1. Data collection and refinement statistics for UPF2 domains.

Identified peptide	Phosphorylated residue	Score
LCN**S**LEESIR	S992	188.85
MVE**S**AVIFR	S886	179.75
DSMTEGENIEEDEEEEEGGAE**T**EEQ**S**GNE**S**EVNE	T1042 or S1046 or S1050	136.66

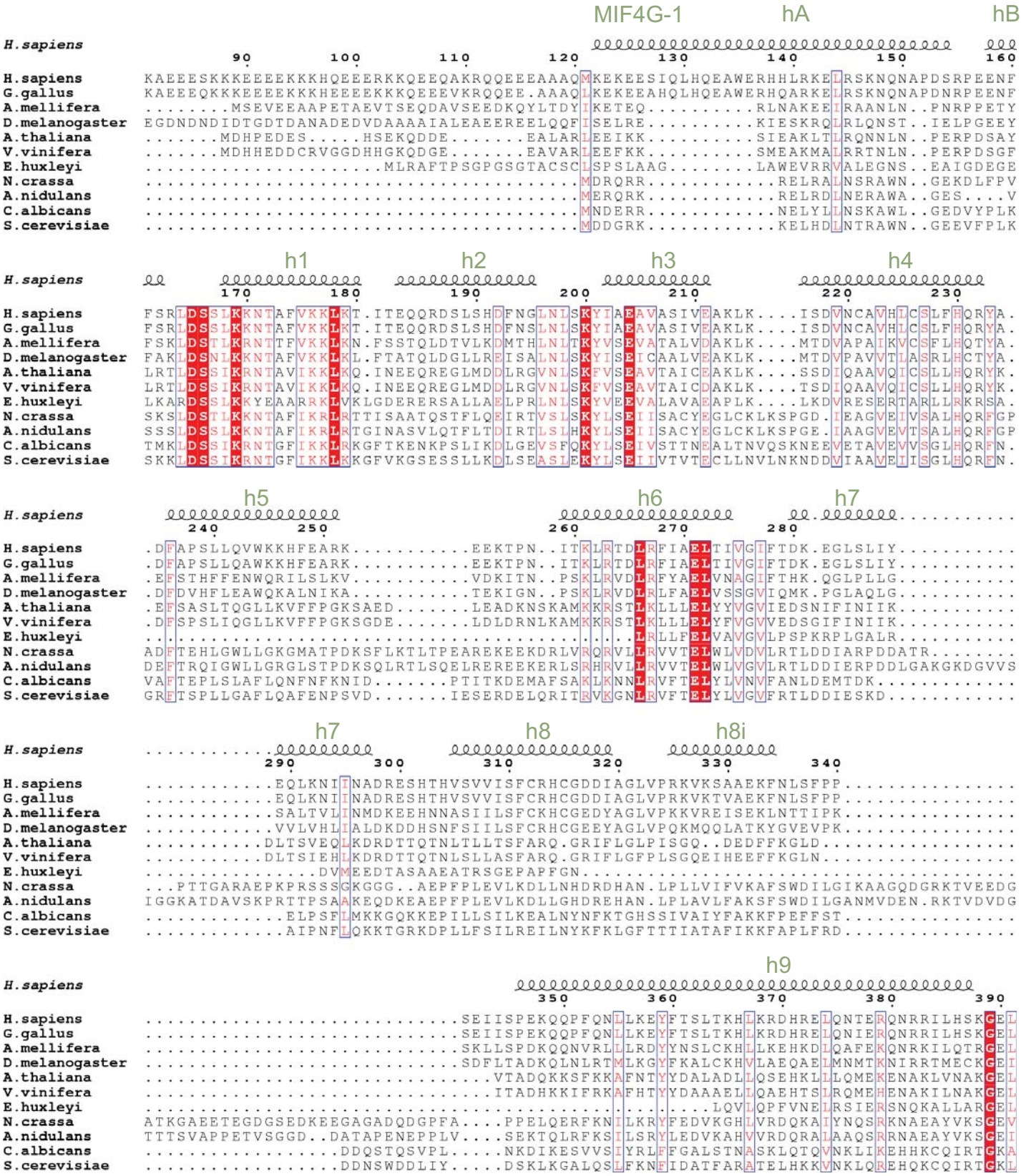
Supplementary Table 2. Identification of the SMG-1 *in vitro* phosphorylation sites of UPF2 by mass spectrometry (MS). Based on MS, only one site is phosphorylated in the third peptide.

	eIF4G	MIF4Gdb	MIF4G-1	MIF4G-2	MIF4G-3	CBP80
eIF4G (1HU3, 744-986)	-					
MIF4Gdb (2I2O, 7-217)	2.9	-				
MIF4G-1 (121-429)	2.9	3.3	-			
MIF4G-2 (458-757)	3.8	3.4	3.5	-		
MIF4G-3 (1UW4, 768-983)	3.6	3.7	3.7	2.5	-	
CBP80 (1H6K, 26-244)	3.5	3.2	4.0	2.5	2.0	-

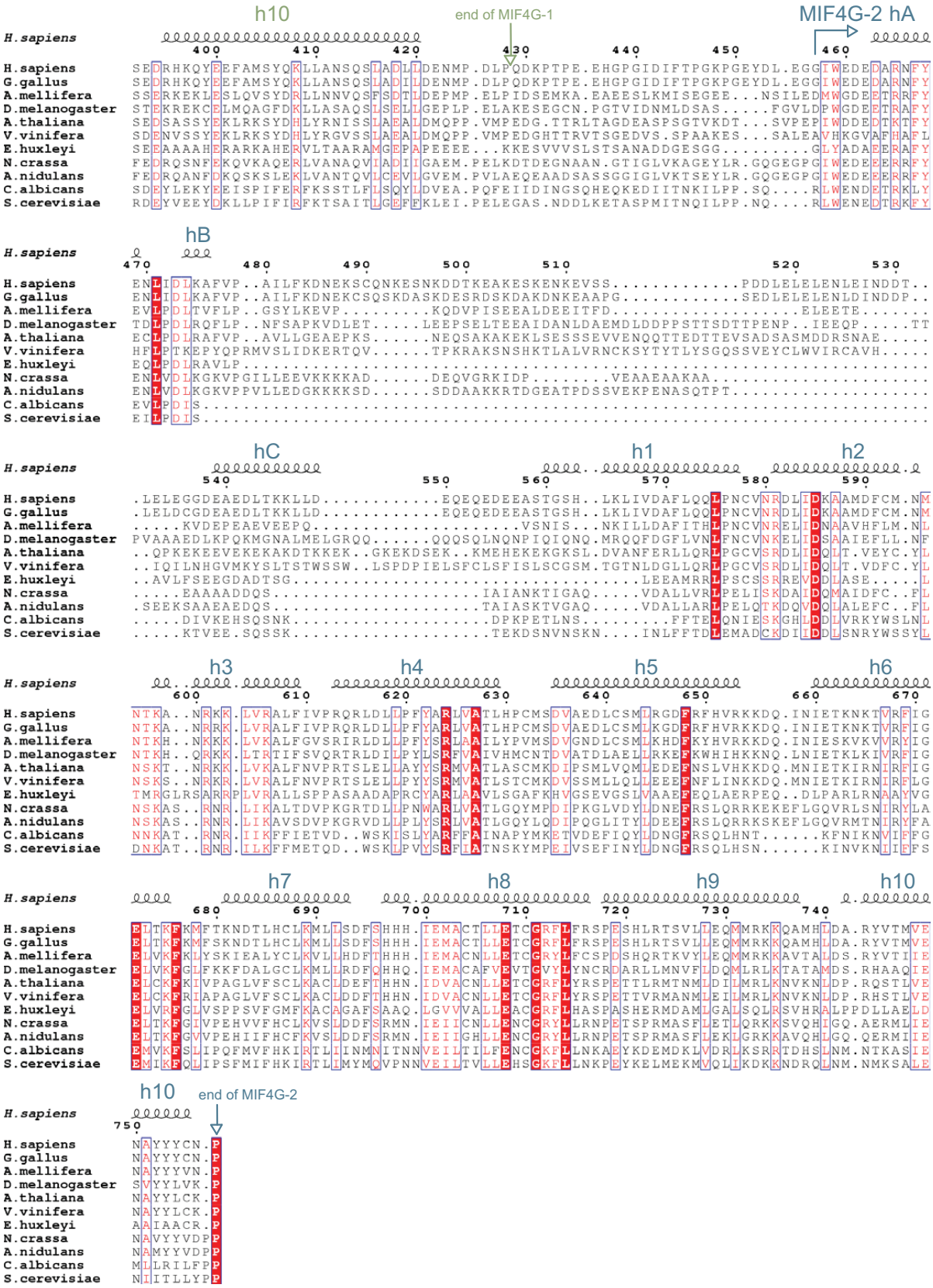
Supplementary Table 3. RMSD (Å) of Cα position for the superposition of MIF4G domains of published structures and UPF2 MIF4G-1 and 2. PDB accession codes and domain boundaries used for superposition are indicated in brackets. The structures were superposed with DaliLite (http://www.ebi.ac.uk/Tools/structure/dalilite).



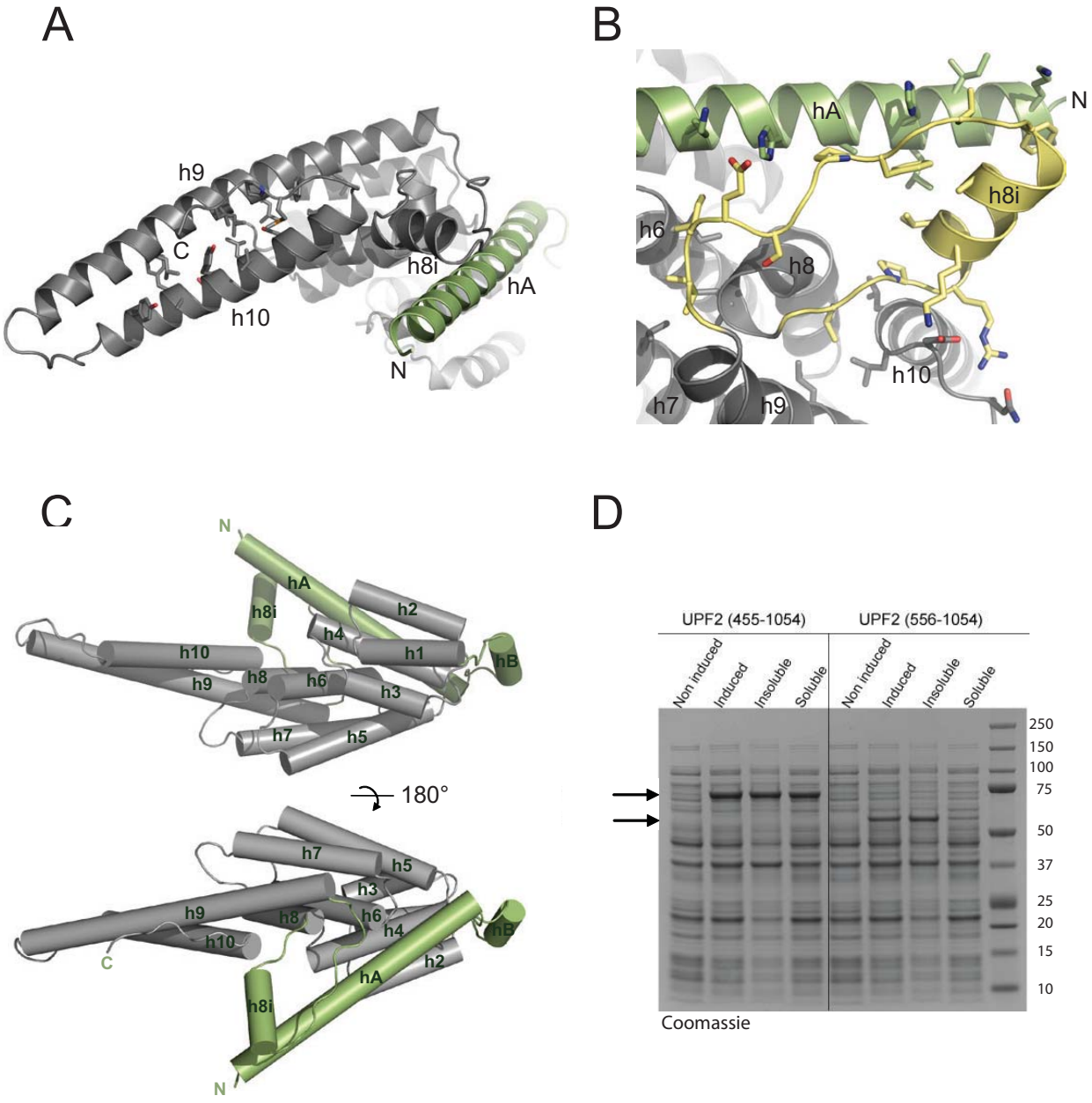
Supplementary Figure S1. A. UPF2(121-1031) limited proteolysis time course and results of N-terminal sequencing on stable bands. UPF2(121-1031) was mixed with trypsin in 1000:1 weight ratio and the reaction stopped at different time points and loaded on SDS-PAGE gel. Stable bands were excised from the gel and analysed by N-terminal sequencing. The N-terminal four residues GAMG are left on the protein from the plasmid used for expression after TEV cleavage. **B**. Schematic representation of UPF2 (rectangle) as in figure 1A. UPF2 constructs used in this study are shown as lines.



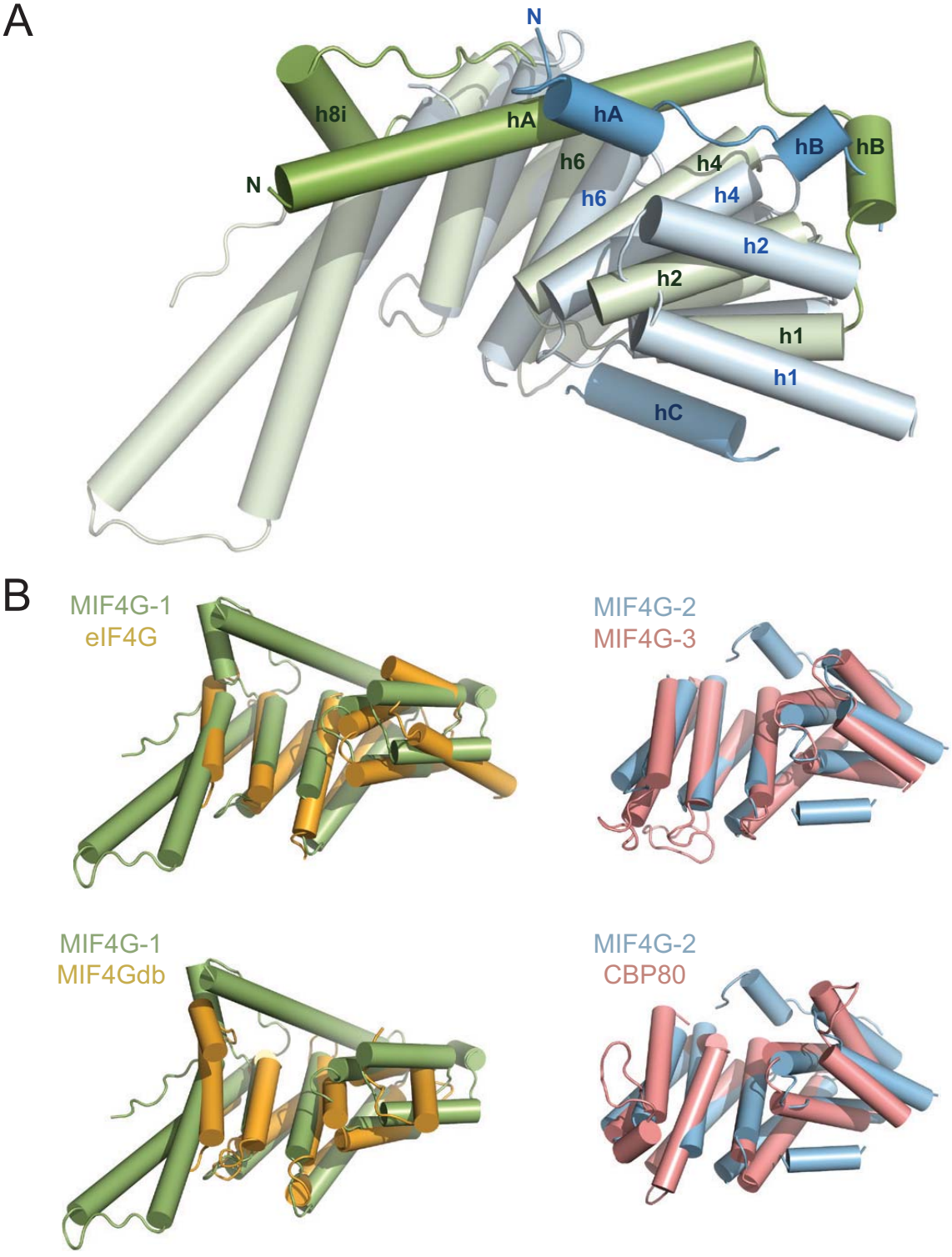
Supplementary Figure S2. (continues in the next page)



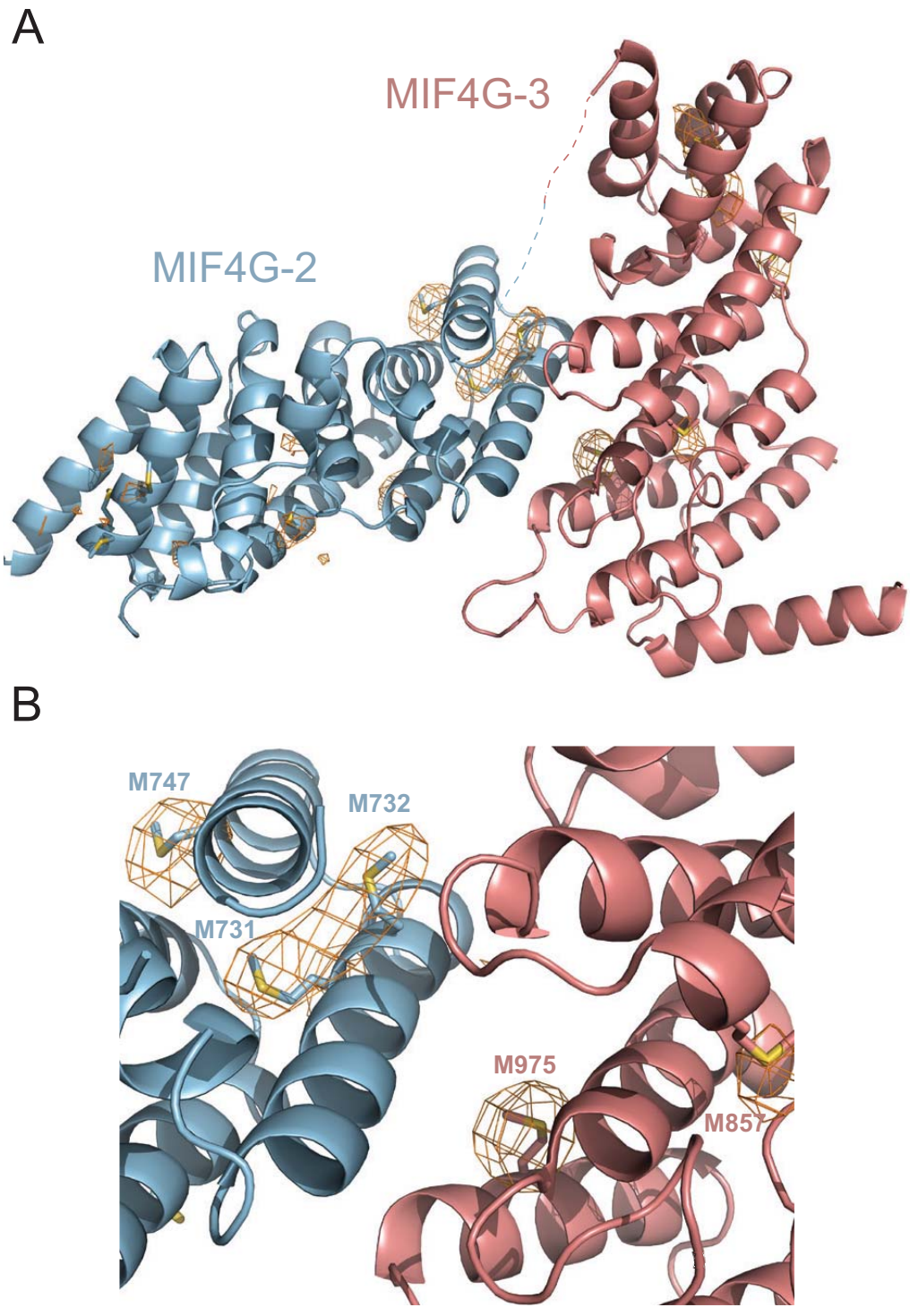
Supplementary Figure S2. Sequence alignment of UPF2 MIF4G domains 1 and 2 of representative UPF2 proteins from yeast to human (Accession numbers. Hs: AAG60689; Gg: XP_004937542; Am: XP_003249451; Dm: NP_572434; At: NP_199512; Vv: XP_002275646; Eh: EOD19187; Nc: XP_961757.2; An: XP_664299.1 ; Ca: XP_720987.1; Sc: NP_011944.2). Residues with similarity >85% are displayed in red. The secondary structure (all alpha-helices) of human UPF2 MIF4G domains 1 and 2 are indicated in green and blue respectively.



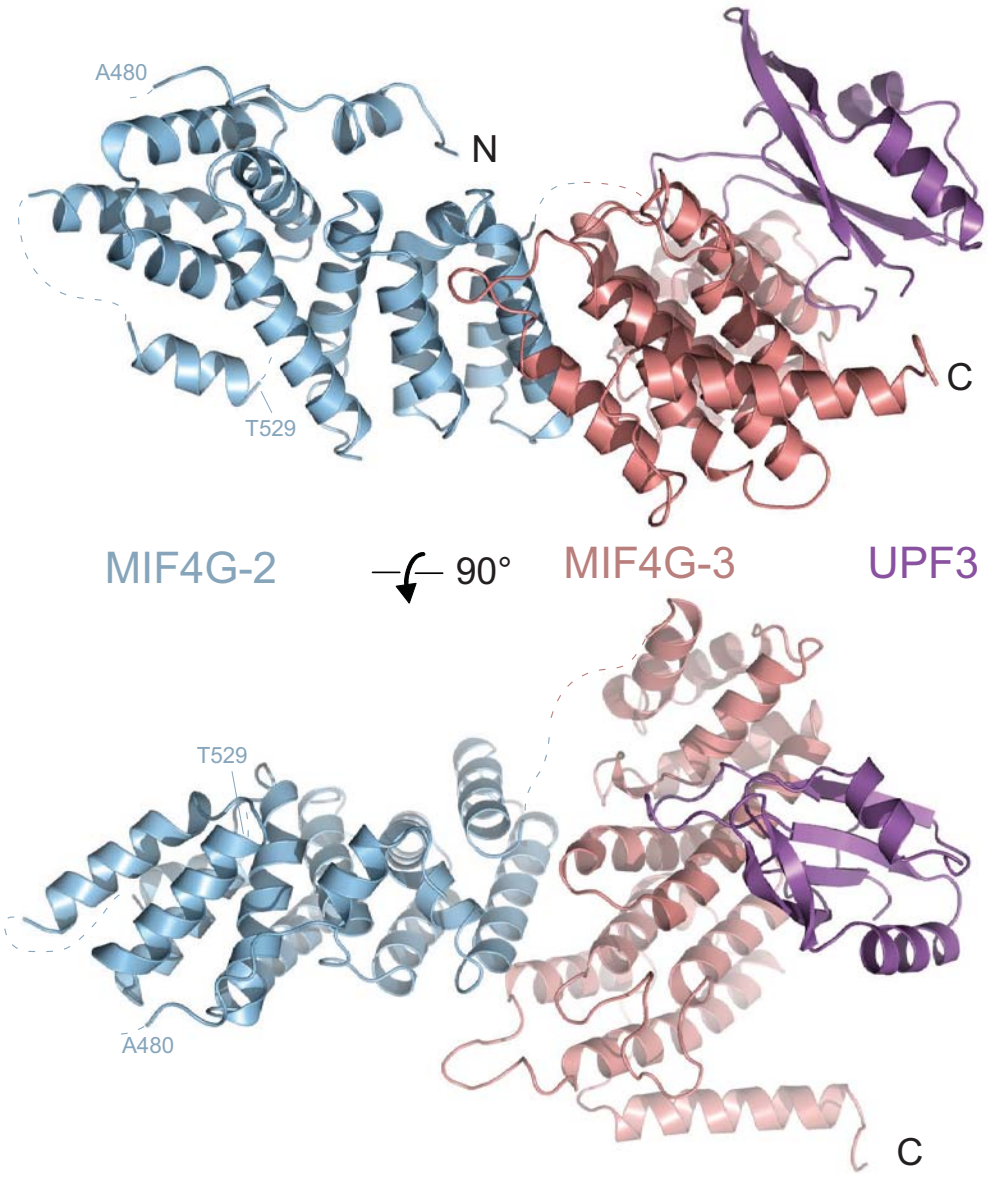
Supplementary Figure S3. A. Helices h9 and h10 of UPF2 MIF4G-1 form an elongated coiled-coil structure protruding away from the domain. Hydrophobic residues involved in helices h9 and h10 packing are depicted as sticks. **B**. Detailed view of the interaction between MIF4G-1 helix h8i with its N- and C-terminal loops and helices h6, h8 and h10. Residues involved in the packing are depicted as sticks. **C**. MIF4G-1 domain represented in the same orientations as in figure 1C. **D**. Coomassie-stained SDS-PAGE of total lysate (non induced and induced), soluble and insoluble fractions of UPF2 MIF4G-2/3 constructs with MIF4G-2 helices hA-hB and disordered linker (constructs 455-1054) and without (556-1054) showing that In the absence of helices hA and hB the solubility of UPF2 MIF4G-2/3 is reduced.



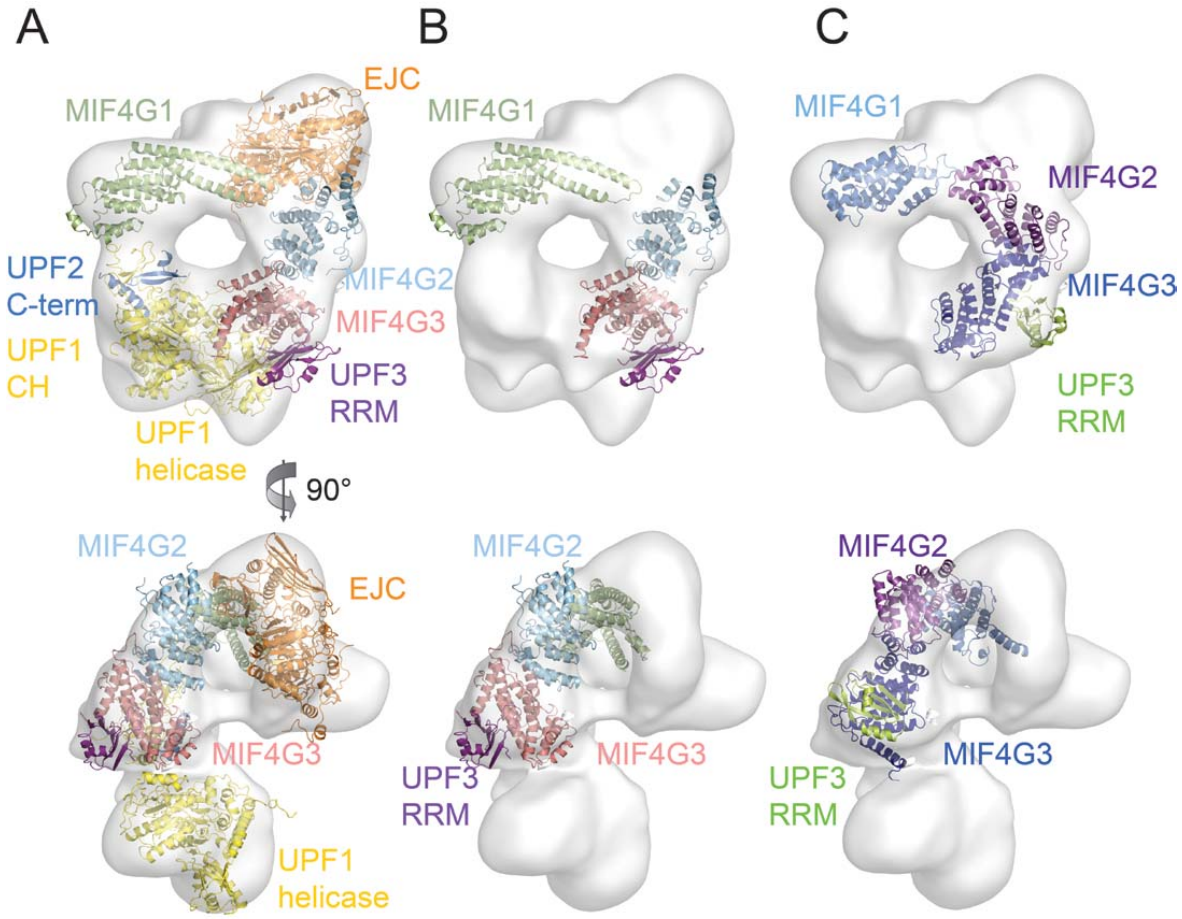
Supplementary Figure S4. A. Superposition of UPF2 MIF4G-1 (light green) and MIF4G-2 (light blue) showing the partial overlap of MIF4G-1 helix hA (green) with MIF4G-2 helix hA (blue). **B**. Superposition of UPF2 MIF4G-1 (green) to eIF4G and MIF4Gdb MIF4G domains (orange); superposition of UPF2 MIF4G-2 (light blue) to UPF2 MIF4G-3 and CBP80 MIF4G domain (pink). Helices are represented as cylinders.



Supplementary Figure S5. A. Anomalous difference map (orange mesh) of selenomethionine substituted UPF2 MIF4G-2/MIF4G-3 domains (contoured at 3.5 σ) indicating the position of the methionine residues (depicted as sticks). **B**. Detailed view of the anomalus difference map in the MIF4G-2/MIF4G-3 interaction region; methionine residues are labeled.



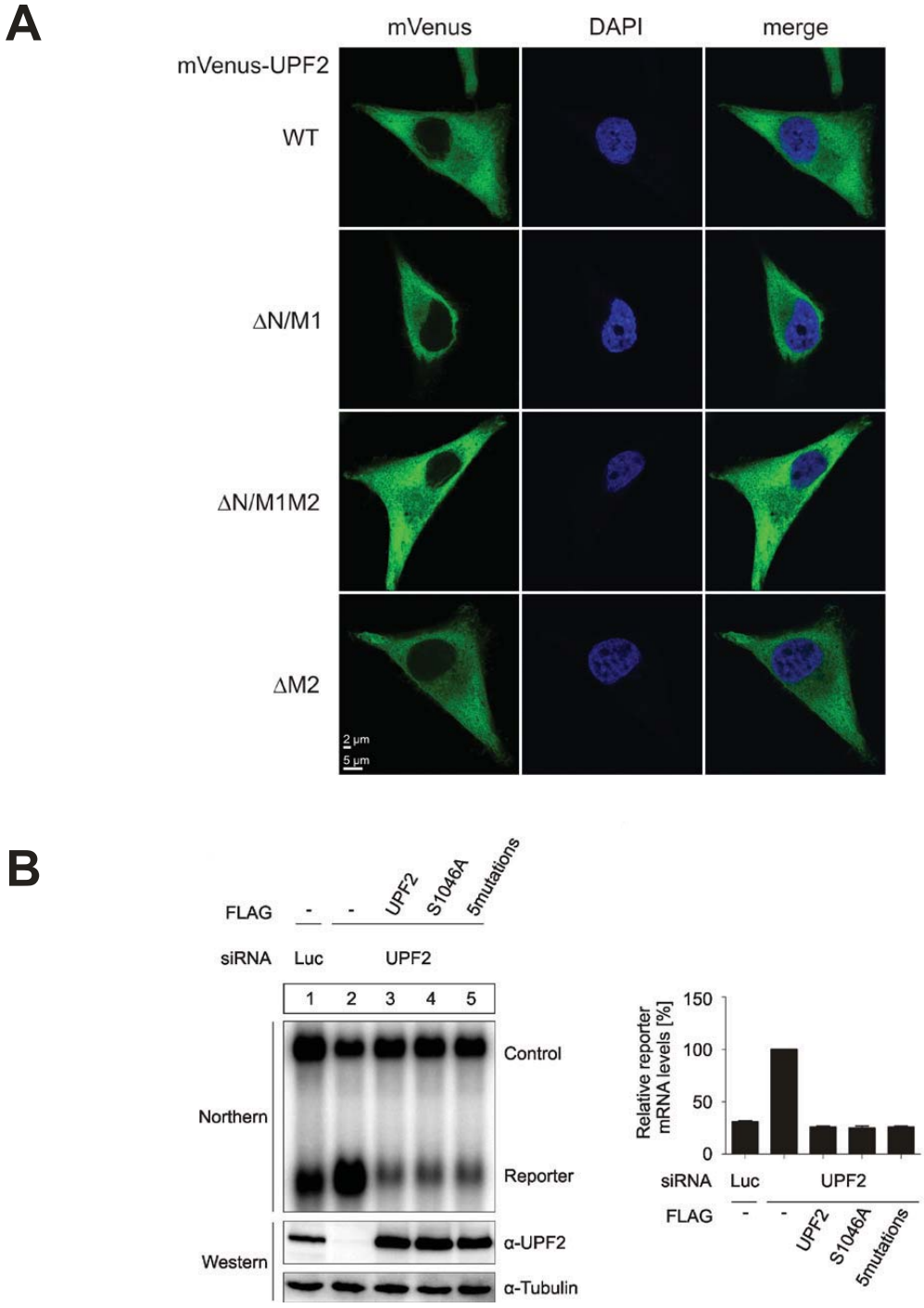
Supplementary Figure S6. Two views of the reconstructed UPF2 MIF4G-2/MIF4G-3 assembly in complex with UPF3 RRM domain. The complex was obtained by superposition of MIF4G-3 in the UPF2-UPF3 complex and in the MIF4G-2/MIF4G-3 assembly.



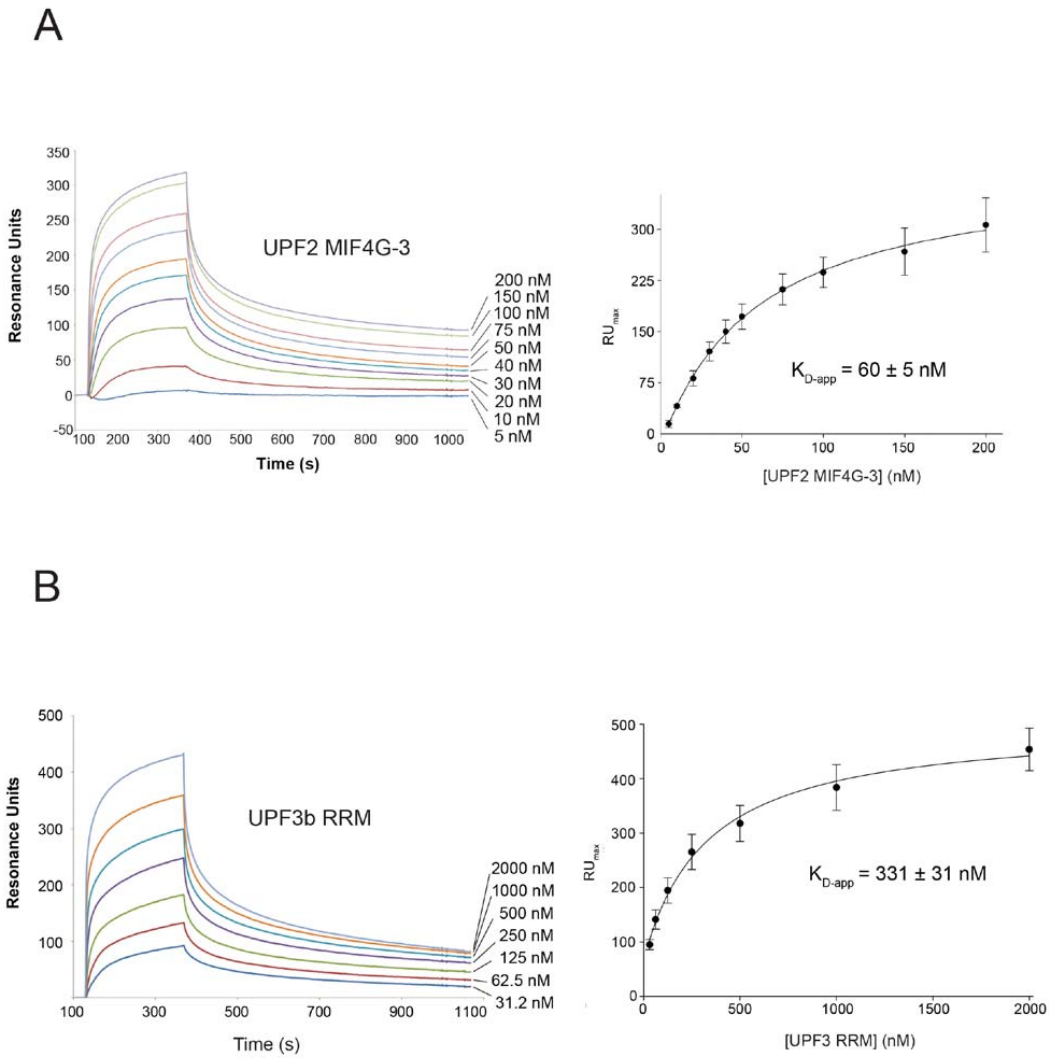
D

Proteins	Correlation Coefficient, published model	Correlation Coefficient, new model
UPF2 MIF4G-domain 1	0.51	0.59
UPF2 MIF4G-domain 2&3 — UPF3 RRM	0.45	0.47
UPF1 CH-domain — UPF2 C-terminus	0.54	0.70
UPF1 helicase	0.34	0.34
EJC — UPF3	0.53	0.58

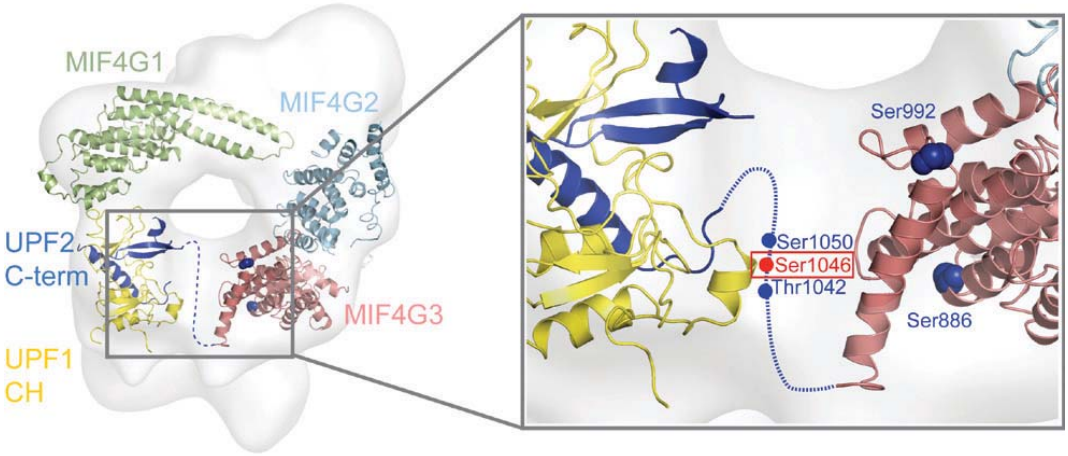
Supplementary Figure S7 A. Fitting of the EJC-UPF3, UPF1 complex with the UPF2 C-terminus, UPF2 MIF4G-1 (this study) and MIF4G-2/MIF4G-3/UPF3 RRM (this study) crystal structures into the cryo-EM reconstruction of the UPF-EJC complex (EMD-2048, Melero et al., 2012). The view in the lower panels is rotated 90 degrees with respect to the upper panel. **B**. The same representation as in A showing only the MIF4G-1 and MIF4G-2/MIF4G-3/UPF3 RRM structures (this study) placed into the EM reconstruction. **C**. Original quasi-atomic model showing the MIF4G-1 homology model, the MIF4G-2 homology model and the MIF4G-3/UPF3 RRM crystal structure placed into the cryo-EM reconstruction (Melero et al., 2012). **D**. Table listing the correlation coefficients obtained by rigid body fitting of UPF domains and the EJC for the published quasi-atomic model and for the new model. The overall quality of the fitting is very similar for the two quasi-atomic models. The correlation coefficients were determined using Chimera.



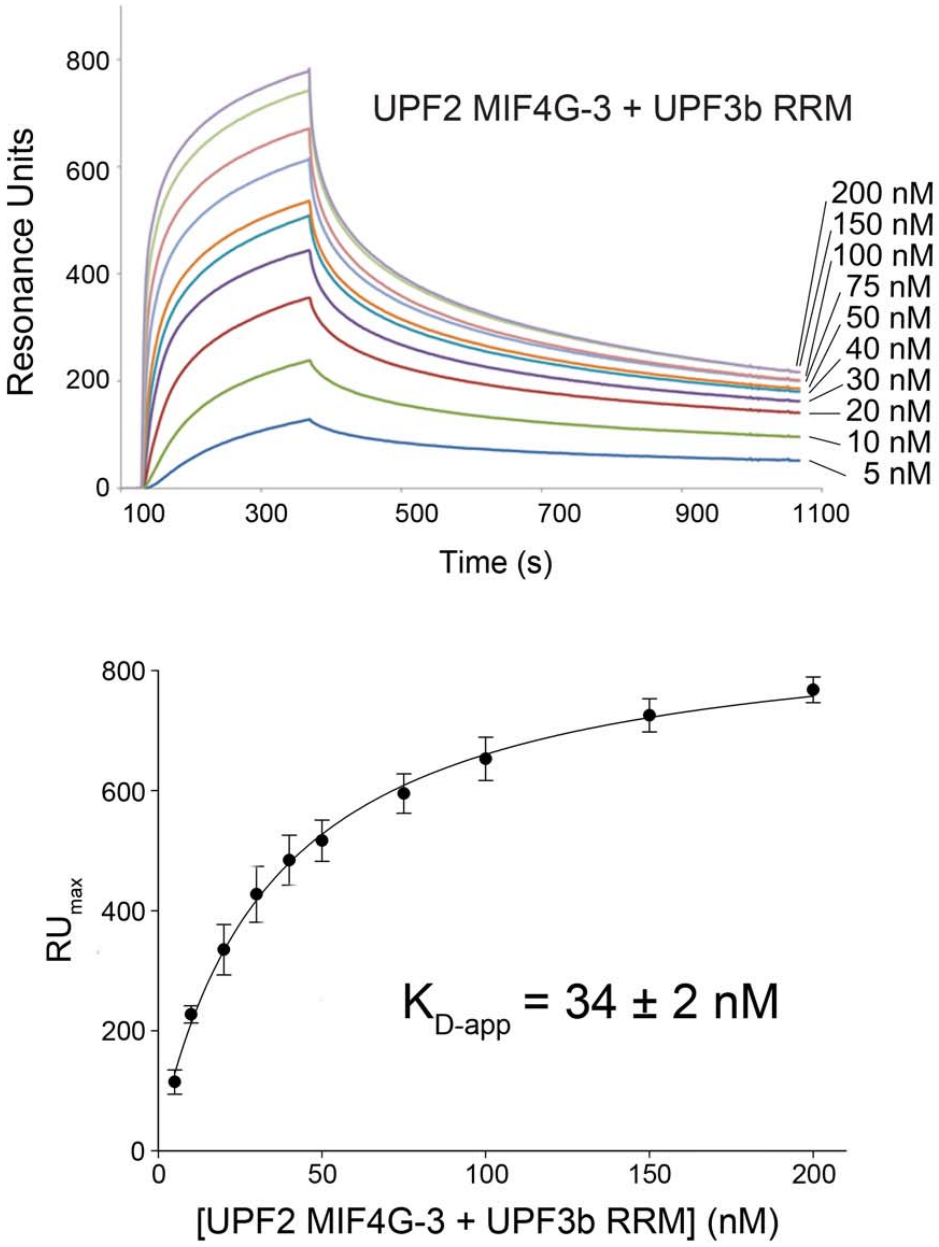
Supplementary Figure S8. A. Subcellular localisation of UPF2 mutants. HeLa cells were transfected with plasmids expressing the indicated mVenus-UPF2-constructs and transferred to coverslips 24 hours afterwards. Cells were fixed 48 hours post-transfection and analysed by confocal microscopy. Images of representative cells from each transfection are shown. Scale bars = 2 and 5mm. All analysed UPF2 mutants localise to the cytoplasm. **B**. Complementation assay performed using UPF2 S1046A mutant and UPF2 carrying all five phosphorylation sites mutations (S886A, S992A, T1042A, S1046A and S1050A). The assay was performed as described in the legend of figure 5B.



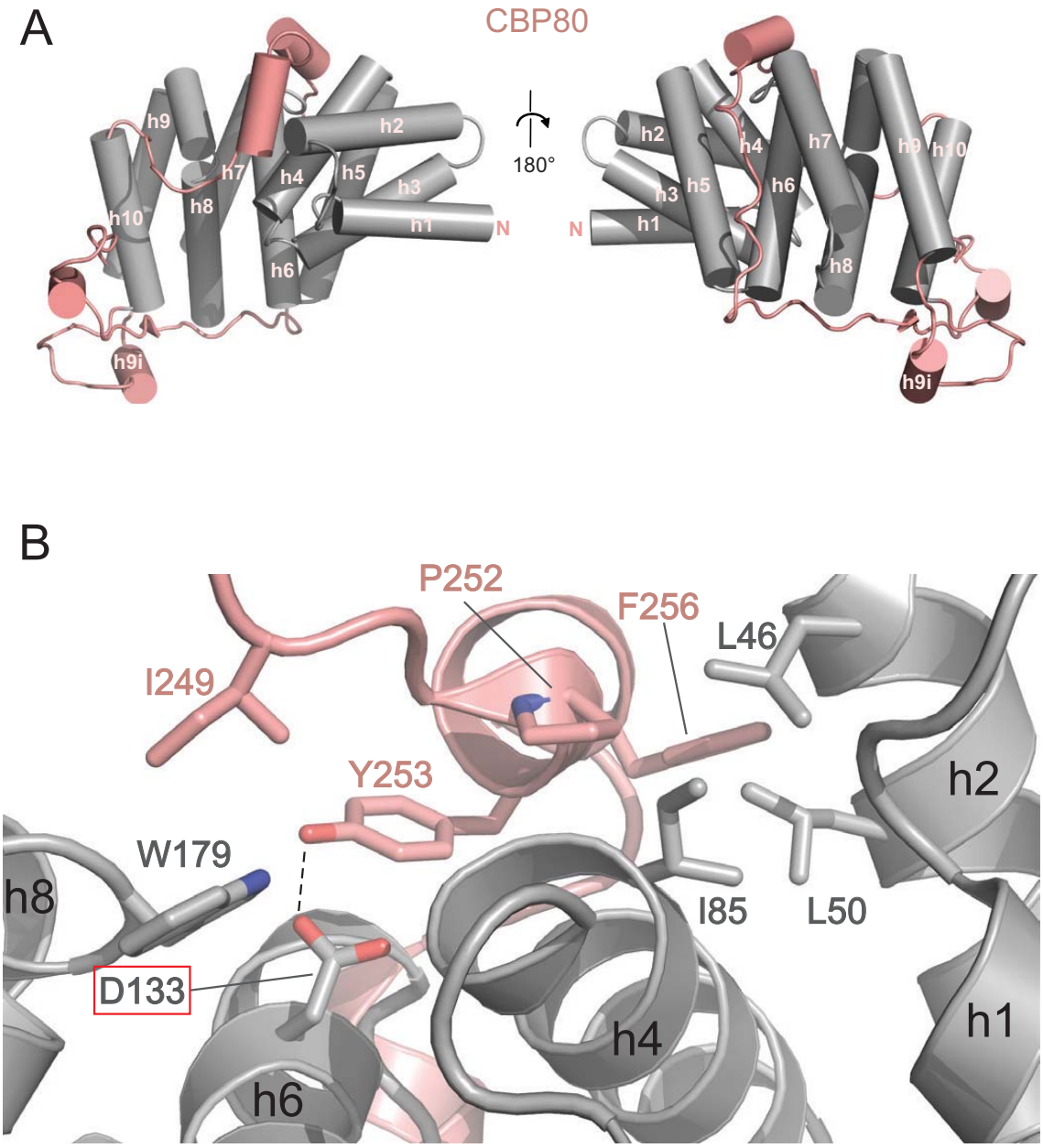
Supplementary Figure S9. Overlay of SPR sensograms of **(A)** UPF2 MIF4G-3 and **(B)** UPF3b RRM interacting with immobilized SMG1. UPF2 MIF4G-3 was injected at concentrations ranging from 5 nM to 200 nM. UPF3 RRM was injected at concentrations ranging from 31.2 nM to 2000 nM. The maximal resonance unit signal was plotted against the injected protein concentration (right panels); the data points result from three independent experiments. The apparent dissociation constants were determined assuming a 1:1 interaction.



Supplementary Figure S10. Location of the SMG1 phosphorylation sites of UPF2 identified by in vitro phosphorylation and mass spectroscopy of UPF2. Phosphorylated serines are depicted as spheres in the quasi-atomic model of the UPF2 – UPF1 CH domain complex which is shown together with the UPF-EJC EM density (Melero et al., 2012). The main in vitro phosphorylation site of UPF2 (Ser1046) is highlighted in red; it is located on a flexible linker between the MIF4G domain 3 and the UPF2 C-terminal domain



Supplementary Figure S11. Overlay of SPR sensograms of the preformed UPF2 MIF4G-3/UPF3b RRM complex interacting with immobilized SMG1. The UPF2/UPF3 complex was injected at concentrations ranging from 5 nM to 200 nM. The maximal resonance unit signal was plotted against the injected protein concentration; the data points result from three independent experiments. The apparent dissociation constant was determined assuming a 1:1 interaction.



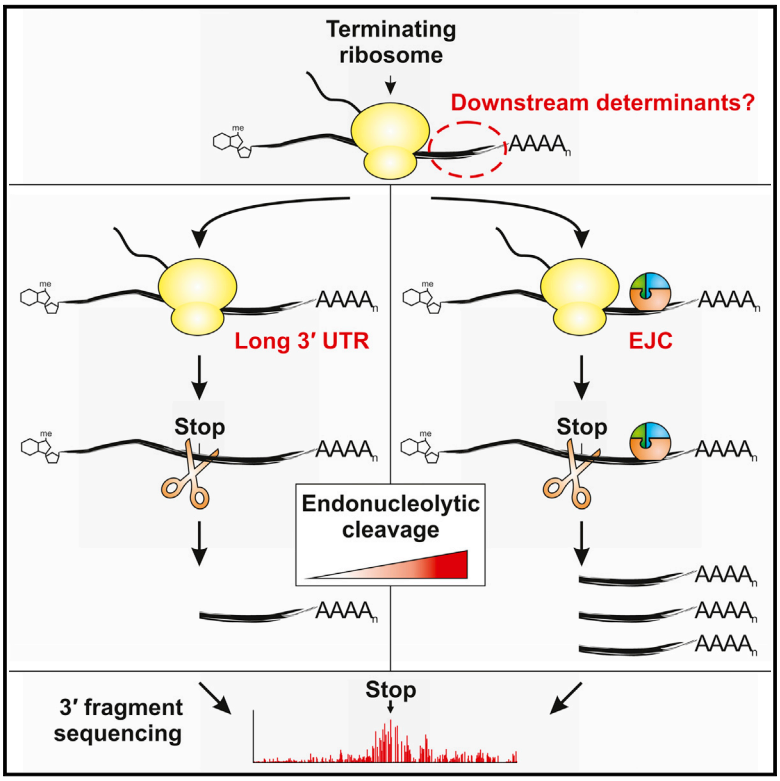
Supplementary Figure S12. A. Structure of human CBP80 MIF4G-1 domain. The ten helices (grey) of the MIF4G domain fold are numbered from one to ten. The extended C-terminal linker of the domain (which contains three helices) and the additional helix (h9i) inserted between h9 and h10 are depicted in pink. **B**. Helix annotation and colour code as in (A). Detailed view of one of the helices of the extended C-terminal linker which interacts tightly with helices h2, h4, h6 and h8 of the core domain. Amino acids involved in the interaction are shown in stick representation. Notably, aspartic acid (D133) of the conserved FIGEL signature motif (FLSDL in CBP80) of MIF4G domains in helix h6 interacts with tyrosine 253 of the C-terminal linker.

Article

Cell Reports

3′ UTR Length and Messenger Ribonucleoprotein Composition Determine Endocleavage Efficiencies at Termination Codons

Graphical Abstract



Highlights

NMD substrates undergo SMG6-mediated endocleavage with different efficiencies

EJCs stimulate endonucleolytic cleavage in a distance-independent manner

The interactions between the EJC and SMG6 or UPF2 are not required for cleavage

Endocleavage of substrate mRNAs occurs at and downstream of the PTC

Authors

Volker Boehm, Nejc Haberman, ..., Jernej Ule, Niels H. Gehring

Correspondence

ngehring@uni-koeln.de

In Brief

Nonsense-mediated mRNA decay (NMD) degrades several classes of mRNAs. Here, Boehm et al. demonstrate that all types of NMD substrates undergo endonucleolytic cleavage, albeit with different cleavage efficiencies and NMD factor requirements. Furthermore, they map the sites of endocleavage using high-throughput sequencing. These data define the molecular characteristics of endocleavage, a major decay pathway for NMD.

Boehm et al., 2014, Cell Reports 9, 555–568
October 23, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.09.012

Cell Reports
Article

3′ UTR Length and Messenger Ribonucleoprotein Composition Determine Endocleavage Efficiencies at Termination Codons

Volker Boehm,[1] Nejc Haberman,[2] Franziska Ottens,[1] Jernej Ule,[2] and Niels H. Gehring[1,*]
[1]Institute for Genetics, University of Cologne, 50674 Cologne, Germany
[2]Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
*Correspondence: ngehring@uni-koeln.de
http://dx.doi.org/10.1016/j.celrep.2014.09.012

SUMMARY

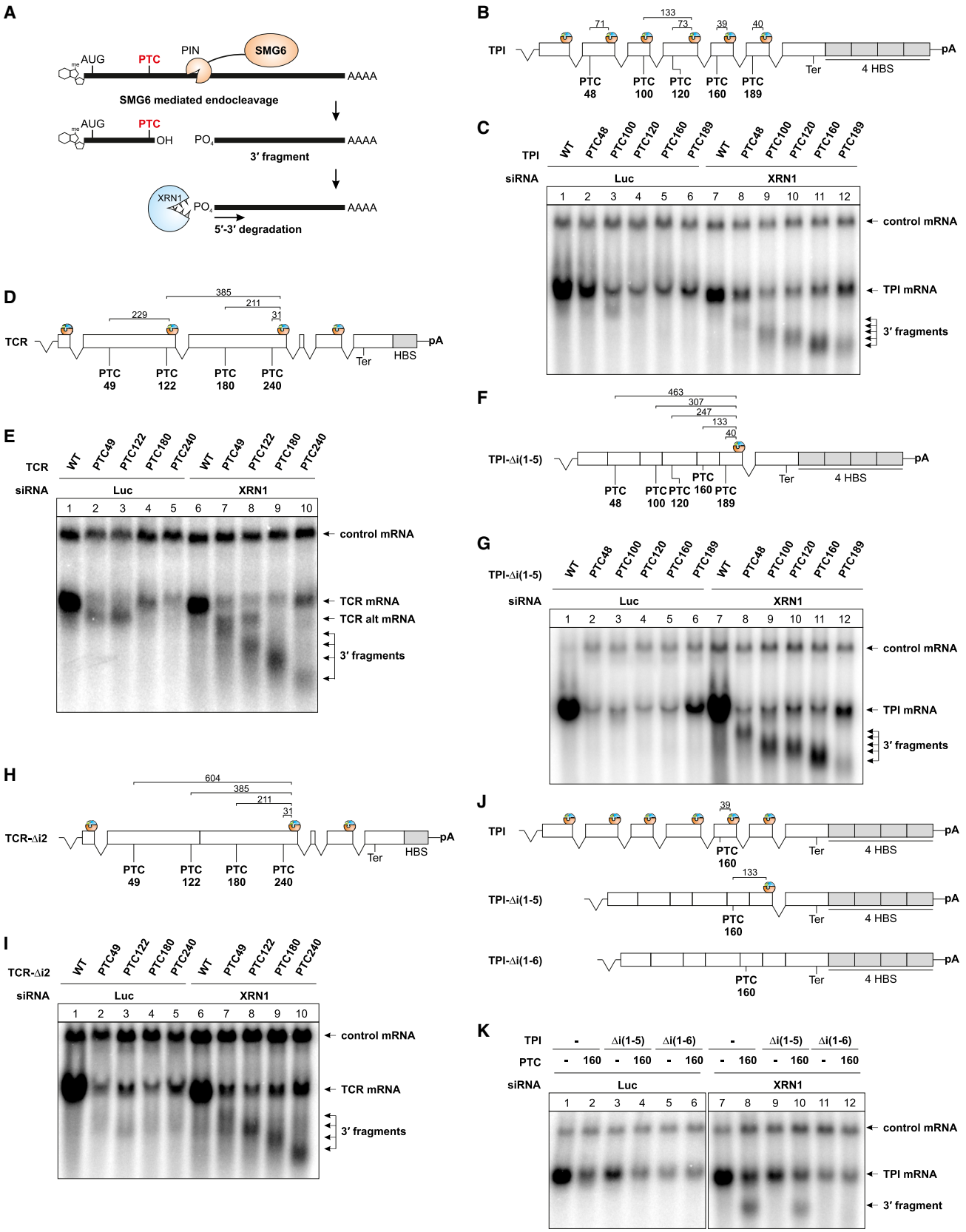
Nonsense-mediated mRNA decay (NMD) degrades different classes of mRNAs, including transcripts with premature termination codons (PTCs). The NMD factor SMG6 initiates degradation of substrate mRNAs by endonucleolytic cleavage. Here, we aim to delineate the cascade of NMD-activating events that culminate in endocleavage. We report that long 3′ UTRs elicit SMG6-mediated endonucleolytic degradation. The presence of an exon-junction complex (EJC) within the 3′ UTR strongly stimulates endocleavage in a distance-independent manner. The interaction of SMG6 with EJCs is not required for endocleavage. Whereas the core NMD component UPF2 supports endonucleolytic decay of long 3′ UTR mRNAs, it is mostly dispensable during EJC-stimulated endocleavage. Using high-throughput sequencing, we map endocleavage positions of different PTC-containing reporter mRNAs and an endogenous NMD substrate to regions directly at and downstream of the termination codon. These results reveal how messenger ribonucleoprotein (mRNP) parameters differentially influence SMG6-executed endonucleolysis and uncover central characteristics of this phenomenon associated with translation termination.

INTRODUCTION

Nonsense-mediated mRNA decay (NMD) is a cellular quality control mechanism that degrades transcripts containing premature termination codons (PTCs) (Chang et al., 2007; Nicholson et al., 2010). NMD exists in all eukaryotic organisms and suppresses the synthesis of C-terminally truncated proteins, which might have inadvertent activities or dominant-negative effects (Holbrook et al., 2004). Whereas a small number of mRNAs carrying disease-causing nonsense mutations have been originally identified to be degraded by NMD (Culbertson, 1999), many so-called endogenous NMD targets are also recognized and degraded by the NMD machinery (Tani et al., 2012). Hence, NMD functions as a general modulator of gene expression and alters the expression levels of ~5% of the transcriptome in eukaryotes (Kervestin and Jacobson, 2012).

In mammalian cells, efficient NMD requires a distance of >50 nt between the termination codon and a downstream exon-exon junction that is marked by an exon-junction complex (EJC) (Neu-Yilik et al., 2001; Thermann et al., 1998). In addition to the canonical EJC-dependent NMD, an alternative EJC-independent NMD targets mRNAs with an aberrant architecture downstream of the termination codon, such as unusually long 3′ UTRs (Bühler et al., 2006; Eberle et al., 2008; Singh et al., 2008). EJC-independent NMD is activated when the ribosome fails to terminate properly, for example, when the interaction of the cytoplasmic poly(A)-binding protein (PABPC1) with the eukaryotic release factors (eRFs) is impaired (Eberle et al., 2008; Fatscher et al., 2014; Singh et al., 2008). As a consequence of any aberrant termination event, the conserved proteins UPF1, UPF2, and UPF3 (UPF3a and UPF3b in humans) assemble into a surveillance complex onto the mRNA and initiate its degradation. According to current models of NMD, UPF1 interacts with the release factors and with the kinase SMG1 upon translation termination, forming the so-called SURF complex (Kashima et al., 2006). In the presence of a downstream EJC, UPF1 is phosphorylated by SMG1 within its extended N- and C-terminal regions (Kashima et al., 2006), which function as docking sites for SMG6 and the SMG5-SMG7 heterodimer (Okada-Katsuhata et al., 2012). SMG5, SMG6, and SMG7 are homologous proteins that interact with phosphorylated UPF1 via their 14-3-3-like domains (Fukuhara et al., 2005; Okada-Katsuhata et al., 2012). An additional phosphorylation-independent interaction between SMG6 and UPF1 has been recently identified (Chakrabarti et al., 2014; Nicholson et al., 2014). SMG6 also contains two N-terminal exon-junction-complex-binding motifs (EBMs), which mediate its direct interaction with EJCs (Kashima et al., 2010).

The degradation of NMD targets is accomplished via redundant pathways that include endonucleolytic cleavage and exonucleolytic degradation (Mühlemann and Lykke-Andersen, 2010). The exonucleolytic decay is primarily coordinated by the SMG5-SMG7 heterodimer (Loh et al., 2013), whereas the endonucleolytic cleavage of NMD targets is mediated by the C-terminal PilT N terminus (PIN) domain of SMG6 (Eberle et al., 2009;



(legend on next page)

Glavan et al., 2006). Endocleavage occurs in the vicinity of the termination codon, and the 5′ and 3′ mRNA fragments are exonucleolytically degraded by the RNA exosome and XRN1, respectively (Eberle et al., 2009; Gatfield and Izaurralde, 2004; Huntzinger et al., 2008).

In view of the many different degradation pathways that are used by the NMD machinery, the rules by which individual NMD targets are selected for a particular degradation pathway remain elusive. Particularly, the incidence of endocleavage by SMG6 has not been systematically studied and the role of *trans*-acting factors is not well defined. Hence, central decisions during the degradation of NMD substrates are not yet fully understood.

In this work, we have analyzed the 3′ fragments (i.e., decay intermediates) of different NMD substrates in cultured human cells. We find that EJCs stimulate the generation of 3′ fragments but do not influence the site of endocleavage. Tethering of individual EJC components and NMD factors recapitulates the stimulation of endocleavage by EJCs and enables the analysis of isolated proteins and their domains. Furthermore, we show that all classes of NMD substrates, including mRNAs with long 3′ UTRs, are degraded via SMG6-dependent endocleavage, suggesting that a common mechanism of SMG6 recruitment underlies EJC-dependent and independent NMD. Using biochemical and high-throughput methods, we analyzed 3′ fragments with nucleotide resolution to show that endocleavage is determined by the position of the termination codon in different NMD substrates.

RESULTS

Endocleavage Is Determined by the Position of the Termination Codon in Multiexon mRNAs

The degradation of NMD substrates in human cells involves SMG6-mediated endonucleolytic cleavage, which generates two mRNA fragments (Figure 1A). Knockdown of the exoribonuclease XRN1 stabilizes 3′ fragments and thereby enables their detection. In this work, we analyze 3′ fragments in XRN1-depleted HeLa cells as an explicit readout for the endonucleolytic activity of NMD.

Previously, the sites of endocleavage during human NMD were mapped to positions in close vicinity of PTCs within short reporter mRNAs (Eberle et al., 2009). In order to analyze endocleavage of complex reporter mRNAs, we first established the detection of 3′ fragments with the well-studied NMD substrate β-globin PTC39 mRNA (Figure S1A; Thermann et al., 1998). After small interfering RNA (siRNA)-mediated depletion of XRN1 (Figure S1B), a specific band for the 3′ fragment was detectable by northern blotting with a probe that hybridizes to the 3′ end of the β-globin mRNA (Figure S1C). Next, we set out to systematically investigate the 3′ fragments of the seven exon triosephosphate isomerase (TPI) reporter mRNA with PTCs located in different exons and at varying distances to the next EJC. Considering the specific and sensitive detection of the β-globin 3′ fragment, we included multiple copies of a short β-globin 3′ UTR sequence (HBS) in the 3′ UTR of the TPI expression construct to facilitate the analysis of TPI 3′ fragments. This modification enabled the detection of the TPI mRNA with a heterologous 3′ probe and enhanced the sensitivity of the northern blot (Figures S1D and S1E). Because stronger 3′ fragment signals were observed with TPI-PTC reporters containing a quadruple HBS cassette (TPI-4H), this cassette was used in all subsequent TPI reporters (Figure S1F).

We studied the expression and 3′ fragment generation of TPI mRNAs with PTCs at positions 48, 100, 120, 160, or 189, which are located in exon 2, 3, 4, 5, and 6, respectively (Figure 1B). Upon XRN1 depletion, we observed abundant 3′ fragments generated from all PTC-containing mRNAs but for the wild-type (WT) mRNA (Figure 1C). Notably, the length of the 3′ fragments inversely correlated with the length of the translated open reading frame up to the PTC, suggesting that the site of endonucleolytic cleavage is primarily determined by the position of the PTC and occurs in close proximity thereof. To confirm our results with an unrelated mRNA, we used the six exon T cell receptor β (TCR) mRNA containing one HBS in the 3′ UTR. Strikingly, we observed a pattern of 3′ fragments for PTCs at positions 49, 122, 180, or 240 located in exons 2 and 3 of the TCR mRNA, which was similar to the 3′ fragments obtained with the TPI mRNA (Figures 1D and 1E). Although we detected slightly different amounts of 3′ fragments for different reporter mRNAs, our results demonstrate that NMD substrates with a complex molecular architecture undergo efficient endocleavage at the termination codon.

Figure 1. Endocleavage of NMD Targets near the PTC Occurs Independently of the Distance to or Position of the Downstream EJC

(A) Schematic overview of SMG6 endocleavage of substrate mRNAs and degradation of 3′ fragments in 5′-3′ direction by XRN1.

(B) Schematic representation of the transfected triosephosphate isomerase (TPI) reporter mRNAs. The position of premature translation termination codons (PTC) are indicated below; the position of exon junction complexes (EJCs) and their distance in nt to the next upstream NMD-activating PTC is indicated above the mRNA. Exons are depicted as white boxes, introns as two connecting black lines, and β-globin 3′ UTR sequences (HBS) as gray boxes. Vector-derived 5′ UTR intron and SV40 poly(A) signal (pA) are indicated. The four repetitions of HBS (4H) serve as binding sites for the heterologous northern probe.

(C) Northern blot of RNA samples extracted from HeLa cells transfected with the indicated siRNAs and reporter constructs. Cotransfected LacZ-H4 served as control mRNA.

(D) Schematic representation of the transfected T cell receptor β (TCR) reporter mRNAs as in (B).

(E) Northern blot analysis as in (C). TCR alt mRNA is an alternative splice variant of the TCR, skipping part of the second exon.

(F) Schematic representation of the transfected TPI reporter mRNAs lacking introns 1–5 as in (B).

(G) Northern blot analysis as in (C).

(H) Schematic representation of the transfected TCR reporter mRNAs lacking intron 2 as in (B).

(I) Northern blot analysis as in (C).

(J) Schematic representation of the transfected TPI reporter mRNAs as in (B).

(K) Northern blot analysis as in (C).

See also Figure S1.

The PTCs in our reporter mRNAs are located 39–133 (TPI) or 31–385 (TCR) nt upstream of the next downstream EJC. According to the current model of NMD, the interaction of ribosome-associated UPF1 with EJC-bound UPF2-UPF3b initiates the degradation of the mRNA (Kashima et al., 2006). Because the bridging between UPF1 and the EJC is believed to be critical for the activation of NMD, we aimed to disturb this interaction by increasing the distance between the PTC and the EJC. To this end, we deleted all but the last intron of the TPI reporter (TPI-Δi(1-5)), which increases the distance between PTC and EJC up to a maximum of ~460 nt for the PTC48 construct (Figure 1F). Notably, despite the increased spacing of PTC and EJC, the average position of endocleavage remained unaltered when we compared the different TPI-Δi(1-5) mRNAs to the corresponding TPI mRNAs with all introns (compare Figures 1C and 1G). Hence, the position of endocleavage does not change when the first EJC downstream of the PTC is moved farther away. To verify this finding, we deleted the second intron of the TCR construct (TCR-Δi2) to create an even larger PTC-EJC distance, which places PTC49 ~600 nt upstream of the next EJC located on exon 3 (Figure 1H). Again, the average cleavage position did not change and the 3′ fragments of the TCR-Δi2 mRNAs were comparable in size to those of the matching TCR mRNA (Figure 1I). In summary, our analysis of two different mRNAs with multiple exons indicates that the position of the EJC does not affect the site of endocleavage, which occurs exclusively proximal to the PTC.

EJCs Are Required for Efficient Endocleavage

Whereas most NMD substrates require at least one EJC downstream of the PTC for their efficient degradation, a small subset of mRNAs has been shown to be degraded in an EJC-independent manner with somewhat reduced efficiency (Bühler et al., 2006). In order to test if the presence of a downstream EJC is dispensable for endocleavage, we removed all introns or all but the first intron from the TPI (TPI-Δi(1-6)) or the TCR (TCR-Δi(2-5)) constructs, respectively (Figures S1G and S1H). Owing to the lack of introns, no EJCs are present downstream of the PTCs in both mRNAs. Although we did not detect a distinct 3′ fragment pattern for both reporters upon XRN1 depletion (Figures S1I and S1J), minor amounts of 3′ fragments appear to be generated at PTCs 100 and 160 of the TPI mRNA, indicating that minimal levels of endocleavage can also occur in the absence of downstream EJCs (Figure S1I). No 3′ fragments were visible for the PTC48 mRNA, which appears to undergo very little or no endocleavage in the absence of splicing. The expression levels of the PTC-containing mRNAs were comparable to that of the WT mRNA, confirming that NMD is not efficiently initiated when all downstream EJCs are removed (Figures S1I and S1J; Neu-Yilik et al., 2001). Because the presence of a downstream EJC appears to be required for efficient endocleavage, we decided to directly compare the endocleavage activity between mRNAs with and without EJCs. To this end, we studied the amount of 3′ fragments generated from three different PTC160-containing mRNAs (TPI, TPI-Δi(1-5), and TPI-Δi(1-6); Figure 1J). Whereas both intron-containing reporters produce strong 3′ fragments of the same size, we did not detect clear 3′ fragments for the intronless construct (Figure 1K). Notably, the expression levels of the reporter mRNAs were increased when at least one intron was present within the open reading frame. We speculate that the lack of EJCs within a large region of the open reading frame may be responsible for the decreased expression levels, although all reporter mRNAs contain an intron in their 5′ UTR and are therefore expected to profit at least partially from the expression-enhancing effects of splicing and EJC deposition (Nott et al., 2004). Together, our data indicate that EJCs support endocleavage at PTCs and directly connect the efficiencies of endocleavage and NMD.

Long 3′ UTRs Trigger Endocleavage at the Termination Codon

Several mRNAs with long 3′ UTRs are regulated by NMD, despite lacking an EJC downstream of the termination codon (Singh et al., 2008). Although mRNAs with long 3′ UTRs represent an established class of NMD substrates, the role of SMG6-mediated endocleavage in mammalian cells has not been addressed. To this end, we either inserted the full-length (5-FL) 3′ UTR of SMG5, a shortened (5) version thereof, the 3′ UTR of UPF3b (3b), or the coding sequence of GFP (GFP) into the 3′ UTR of the TPI WT construct (Figure 2A). Whereas the mRNAs encoding SMG5 and UPF3b are natural NMD targets due to their long 3′ UTR, the heterologous GFP coding sequence is sufficiently long to activate NMD as well (Singh et al., 2008). Strikingly, we observed 3′ fragments produced from all four reporter mRNAs upon XRN1 depletion. The sizes of these 3′ fragments corresponded to the lengths of the respective 3′ UTRs and hence indicate that endocleavage occurs in the vicinity of the TPI stop codon (Figure 2B). The levels of these 3′ fragments were strongly reduced when XRN1 and SMG6 were codepleted, demonstrating that they originate from SMG6-dependent, bona fide NMD (Figure 2B, lanes 9–12).

Whereas our results suggest that the 3′ fragments of long 3′ UTRs are generated at positions surrounding the TPI termination codon, we cannot exclude that this region is particularly susceptible to endocleavage independently of translation termination. To this end, we generated intronless TPI constructs containing different PTCs and the SMG5 or UPF3b 3′ UTR, respectively (Figure 2C). Thereby, we moved the stop codon to more upstream positions and simultaneously extended the 3′ UTR, while preventing the deposition of EJCs downstream of the PTC. Hence, endocleavage within these mRNAs is activated by the long 3′ UTR and not stimulated by an EJC. In line with our earlier observations, the length of 3′ fragments corresponded to the length of the 3′ UTRs and longer 3′ fragments were detected for mRNAs with earlier PTCs (Figure 2D). In summary, our results show that endocleavage at the termination codon occurs independently of the mechanism of NMD activation and is a common unifying feature of different classes of mammalian NMD substrates.

SMG6 EBMs Are Dispensable for Endocleavage

Thus far, our results demonstrate that the presence of an EJC strongly stimulates endocleavage even over a long distance. This raises the question of how the surveillance complex, which is assembled at the termination codon, recognizes the presence of a downstream EJC. Notably, SMG6 itself contains two

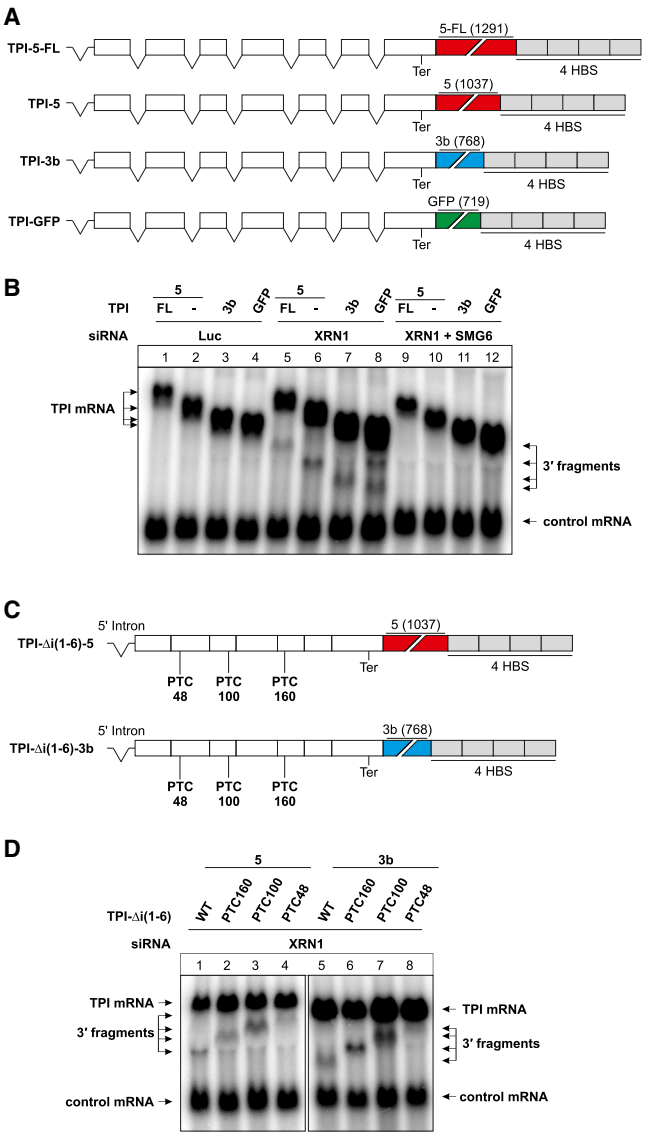


Figure 2. Long 3′ UTRs Trigger SMG6-Mediated Endocleavage near the Termination Codon

(A) Scheme of the long 3′ UTR reporter mRNAs depicted as in Figure 1, with the inserted sequences (SMG5 full length [5-FL] or shortened [5] and UPF3b [3b] 3′ UTR or GFP coding sequence) shown as colored boxes. Length of the inserts is indicated above the mRNA in brackets. EJC positions and poly(A) signal are not shown.

(B) Northern blot analysis of RNA samples derived from HeLa cells as described in Figure 1, with β-globin WT serving as control mRNA.

(C) Schematic representation of the transfected reporter mRNAs as in (A).

(D) Northern blot analysis as in (B).

N-terminal EBMs, which are short peptide sequences that bind directly to EJCs (Kashima et al., 2010). Because the EBMs of SMG6 have been shown to be essential for mammalian NMD (Kashima et al., 2010), we speculated that the EBMs may either promote the formation of a stable degradation complex including the EJC or stimulate the endonuclease activity of SMG6 in an EJC-dependent manner. In order to delineate the role of the SMG6 EBMs during endocleavage, we established a complementation assay for SMG6 in HeLa cells. After siRNA-mediated depletion, the expression of SMG6 is rescued by the transfection of a siRNA-insensitive SMG6 expression plasmid. Consistent with previous reports, the knockdown of SMG6 in XRN1-depleted cells greatly reduced the levels of 3′ fragments generated at all PTCs in the TPI mRNA (Figure 3A). These 3′ fragments are not generated from decapped and XRN1-degraded mRNAs, because we did not detect substantial differences in reporter and 3′ fragment levels when DCP2 and XRN1 were codepleted (Figures S2A–S2C). After SMG6 depletion, the generation of 3′ fragments of TPI-Δi(1-5)-PTC160 could be fully restored by the expression of SMG6 WT, whereas SMG6 with a mutant PIN domain (PIN^{Mut}), lacking the catalytically important triad of aspartate residues in the active site (Eberle et al., 2009; Glavan et al., 2006; Huntzinger et al., 2008), was inactive (Figures 3B–3D). Surprisingly, full recovery of endocleavage was also achieved with a SMG6 mutant that did not contain functional EBMs (EBM^{Mut}; RRP45-47AAA and KKP139-141AEA; Figures 3C and 3D). This effect was independent of the position of the PTC, because similar results were obtained with mRNAs containing PTCs 100 and 189 (Figures S2D and S2E). We observed comparable results in the SMG6 complementation assay with a reporter mRNA that contained a long 3′ UTR (TPI-5). SMG6 WT and SMG6 EBM^{Mut} were able to support 3′ fragment generation, whereas the SMG6 PIN^{Mut} was not (Figures S2F and S2G), showing that the catalytic activity is responsible for the endocleavage at long 3′ UTR mRNAs. To verify that the mutations in the EBM of SMG6 (EBM^{Mut} construct) abolish binding to the EJC, we tested this mutant in binding assays in vitro as purified FLAG-tagged N-terminal fragment of SMG6 (1–207). This fragment contains both EBMs and has previously been shown to be sufficient for binding to EJCs (Kashima et al., 2010). We observed that the WT fragment immunoprecipitated preassembled EJC proteins, whereas the SMG6 EBM^{Mut} and an EBM1+2 deletion mutant failed to interact with the EJC (Figure 3E). To exclude that a residual EJC-binding activity of the EBM^{Mut} SMG6 protein mediates its function in the complementation assays, we performed additional complementation assays with EBM deletion mutants (ΔEBM1, ΔEBM2, and ΔEBM1+2; Figure S2H). Remarkably, also EBM deletion mutants restored reporter and 3′ fragment levels to the full extent (Figures S2I and S2J). Hence, the interaction of the SMG6 EBMs and downstream EJCs is not required for endocleavage of both EJC-dependent and EJC-independent NMD substrates.

UPF2 Is Required for Endocleavage Triggered by Long 3′ UTRs

The dispensability of the SMG6 EBMs during endocleavage prompted us to explore the role of the canonical bridging factor UPF2 during the generation of 3′ fragments. As described above, UPF2 is thought to link the EJC-bound UPF3b to the ribosome-bound UPF1. The depletion of UPF2 has been shown to upregulate certain NMD substrates, but an UPF2-independent NMD pathway has also been described (Gehring et al., 2005). To study the function of UPF2, we used a knockdown protocol that we established previously to analyze the function of different UPF2 domains (Clerici et al., 2014). Surprisingly, 3′ fragments were still

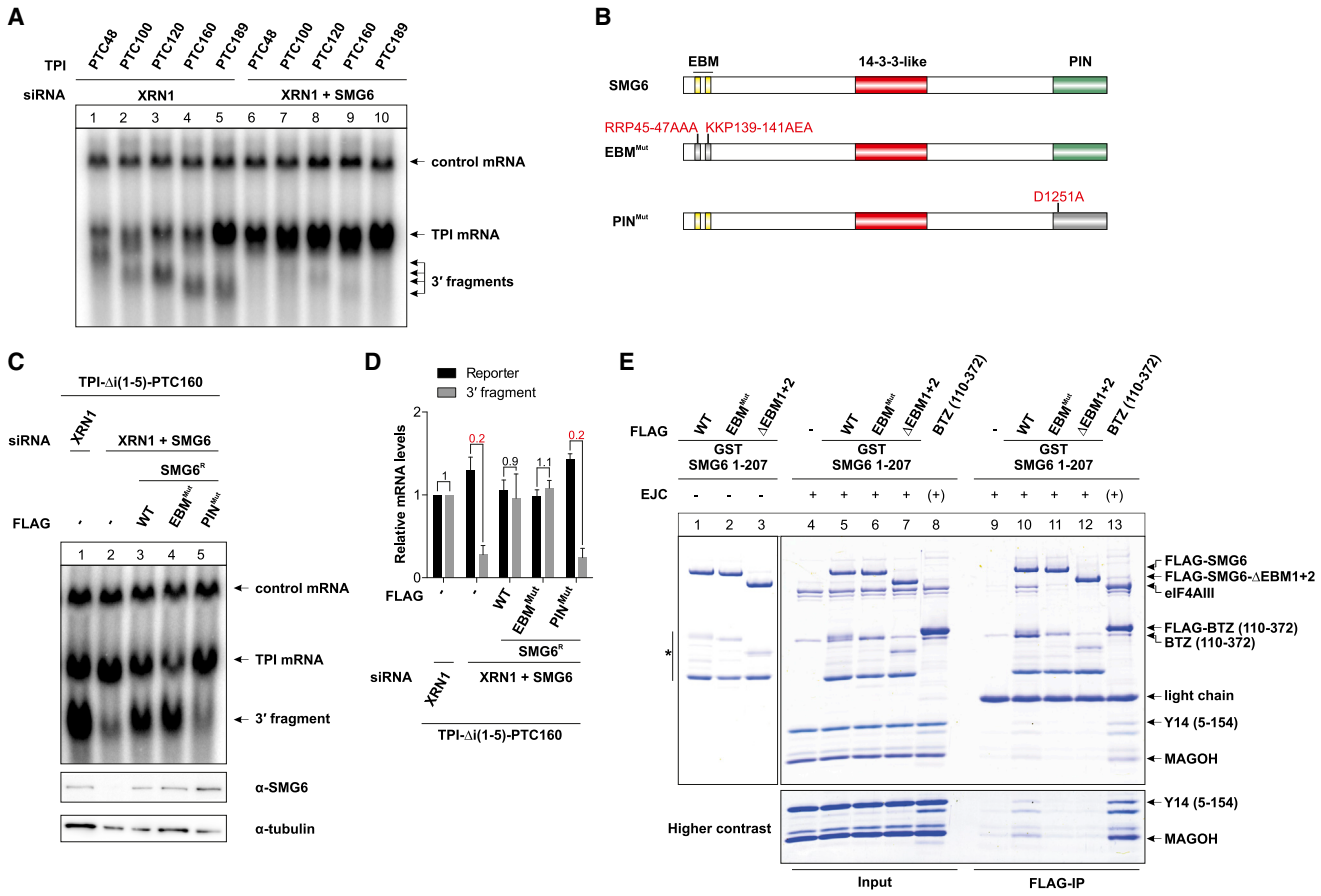


Figure 3. Functional SMG6 PIN Domain Is Required for Endocleavage of Both Types of NMD Substrates

(A) Total RNA of HeLa cells, transfected with the indicated siRNAs and reporter plasmids, was analyzed by northern blotting.

(B) SMG6 protein architecture is depicted schematically, highlighting the N-terminal EJC-binding motifs (EBMs), the central 14-3-3-like domain, and the C-terminal PIN domain. Mutant constructs with the respective mutated residues and their positions are indicated.

(C and D) Complementation assay of SMG6-depleted cells using TPI-Δi(1-5)-PTC160 as mRNA reporter. Knockdown and rescue with siRNA-resistant SMG6 constructs is shown by western blots, using tubulin as loading control (C, lower panels). Northern blotting was performed of RNA extracted from the same cells used for western blot analysis (C, upper panel). Mean values of reporter and 3′ fragment signal ± SD (n = 3) were quantified and normalized to the XRN1 + Luc control knockdown (D). The ratio of 3′ fragment to reporter mRNA levels is indicated above the bars.

(E) FLAG-immunoprecipitation of recombinant SMG6 fragments. FLAG-tagged glutathione S-transferase (GST)-SMG6 1-207 WT or mutant proteins were used to coprecipitate preassembled EJC complexes containing eIF4A3, Y14 (5-154), MAGOH, and BTZ (110-372) (denoted by +). FLAG-tagged BTZ (110-372) was used as a control; untagged BTZ (110-372) was omitted from this reaction, denoted by (+).

See also Figure S2.

generated at PTCs 100, 120, 160, and 189 in UPF2-depleted cells (Figure 4A). In contrast, the 3′ fragment resulting from endocleavage at PTC48 was clearly reduced by the knockdown of UPF2 (Figure 4B). Hence, our data indicate that some EJC-dependent endocleavage events (PTC48) require normal levels of UPF2, whereas others (PTCs 100–189) do not. However, the molecular basis for the differential response to the depletion of UPF2 was not further investigated.

In order to investigate the possibility that SMG6 and UPF2 represent redundant bridging factors, we performed SMG6 complementation assays in XRN1-UPF2-SMG6 triple-knockdown cells (Figure S3A). 3′ fragment levels were restored by both SMG6 WT and EBM^{Mut} (Figures S3B and S3C), showing that UPF2 and the EBMs of SMG6 do not overlap in function and are neither individually nor in combination required for EJC-dependent endocleavage.

We next investigated whether the endocleavage of mRNAs with long 3′ UTRs was affected by UPF2 depletion. Although EJC-dependent endocleavage at PTCs was not or only mildly decreased in UPF2-depleted cells, 3′ fragments of long 3′ UTRs were reduced to undetectable levels after UPF2 knockdown (Figure 4C). These data reveal a striking difference between the endocleavage of PTC containing mRNAs, which is mainly UPF2 independent, and the UPF2-dependent endocleavage of mRNAs with long 3′ UTR.

To validate the role of UPF2, we investigated the endocleavage at long 3′ UTRs with a complementation assay for UPF2 (Clerici et al., 2014). UPF2 binds to UPF1 via the UPF1-binding

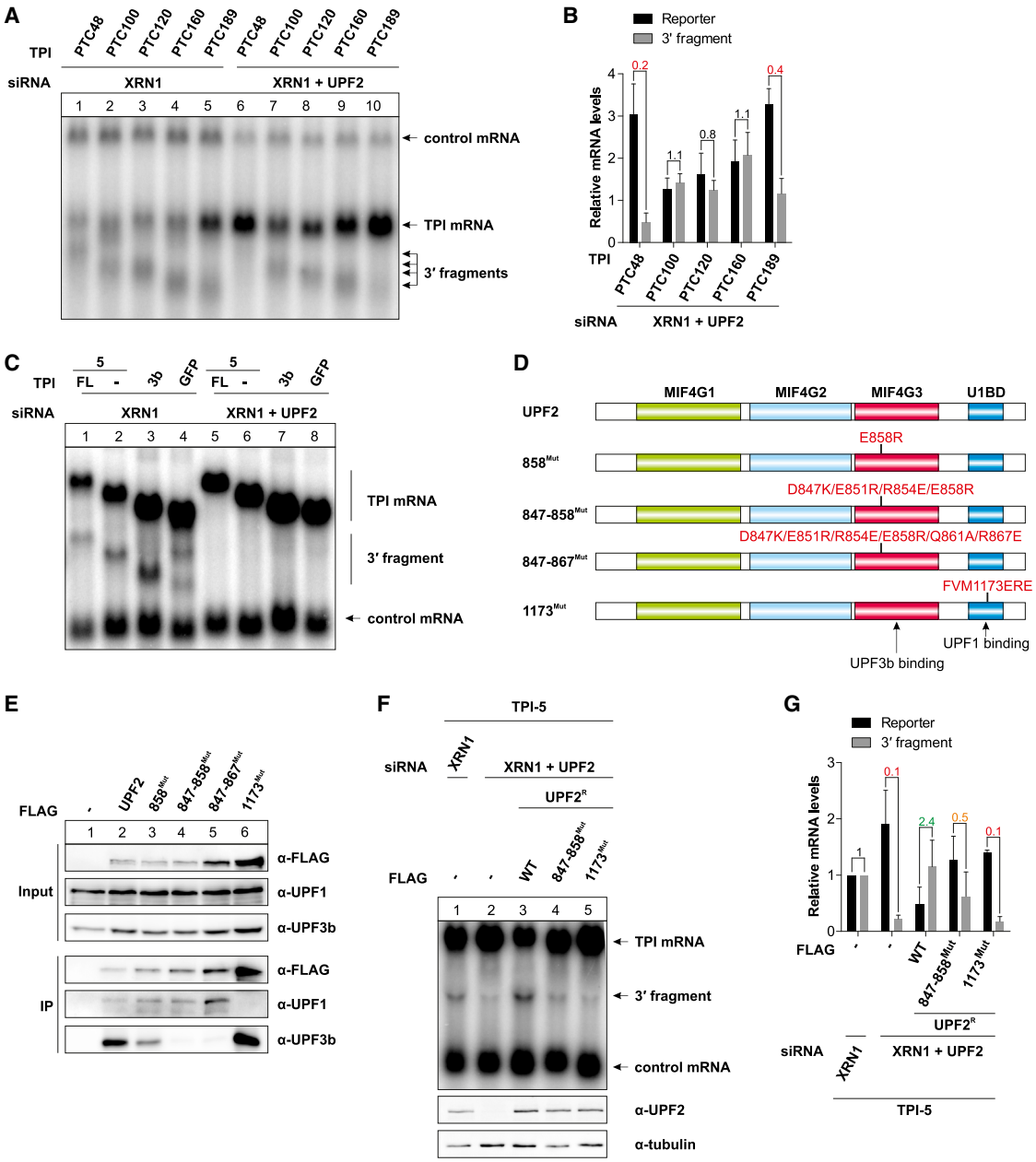


Figure 4. UPF2 and the Interaction with UPF1 Are Required for Endocleavage of Long 3′ UTR mRNAs

(A and B) Transfection with the indicated siRNAs and reporter plasmids was performed in HeLa cells, and total RNA was analyzed by northern blotting. mRNA levels of XRN1 + UPF2 knockdown samples (A) were quantified (mean ± SD; n = 3) and normalized to the respective XRN1 knockdown lanes (B). The ratio of 3′ fragment to reporter mRNA levels is indicated above the bars. LacZ-4H-expressing plasmid was cotransfected as control mRNA.

(C) HeLa cells were transfected with the indicated siRNAs and reporter plasmids, and total RNA was analyzed by northern blotting. β-globin WT-expressing plasmid was cotransfected as control mRNA.

(D) Domain architecture of UPF2 and position of mutations in the MIF4G3- and UPF1-binding domain (U1BD) are shown schematically as in Figure 3B.

(E) Interaction between FLAG-tagged immunoprecipitated UPF2 proteins and endogenous UPF1 and UPF3b in RNase A-treated HeLa cell lysates was analyzed by western blotting.

(F and G) Complementation assay of UPF2 knockdown in HeLa cells. The indicated siRNAs and plasmids were transfected, and TPI reporter mRNA as well as UPF2 protein levels were analyzed by northern and western blotting, respectively. Tubulin serves as a loading control for western blotting. Quantification (mean ± SD; n = 3) of the northern blots is shown in (G), and the ratio of 3′ fragment to reporter mRNA levels is indicated above the bars.

See also Figure S3.

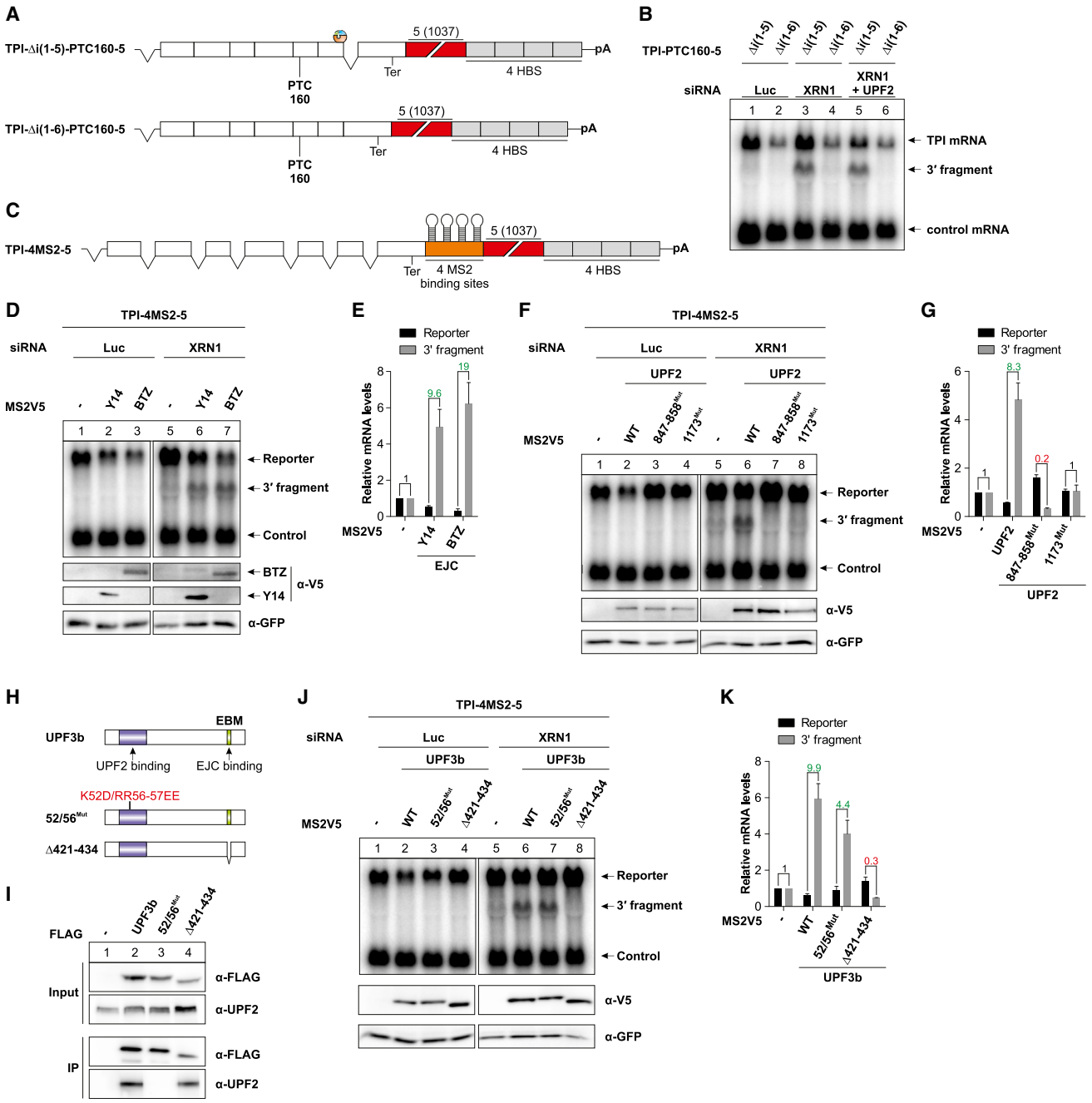


Figure 5. EJC and NMD Factors Enhance NMD and Endocleavage of Long 3′ UTRs

(A) SMG5 3′ UTR-containing TPI-PTC160 reporters with (Δi(1-5)) or without (Δi(1-6)) the last TPI intron are depicted schematically as in Figure 2.

(B) Northern blot of RNA samples extracted from HeLa cells transfected with the two reporters described in (A) in control (Luc), XRN1, and XRN1+UPF2 knockdown conditions. LacZ-4H-expressing vector was cotransfected and served as transfection control.

(C) Scheme of the TPI-4MS2-5-tethering reporter. The four MS2-binding sites allow tethering of MS2 coat protein fusion proteins to the position upstream of the long 3′ UTR but downstream of the stop codon.

(D) HeLa cells were transfected with the indicated siRNAs and plasmids. The effects of tethering the indicated proteins to the TPI-4MS2-5 reporter mRNA were analyzed by northern blotting (top panel). V5-tagged MS2 coat protein served as tethering control. Western blot shows the expression levels of the MS2V5-tagged constructs (lower panel). Cotransfected GFP serves as transfection control.

(E) Mean values ± SD (n = 3) of the mRNA levels in (D) were quantified. The ratio of 3′ fragment to reporter mRNA levels is indicated.

(F) Northern blot analysis as in (D).

(G) Quantification of the mRNA levels in (F) as in (E).

(H) Schematic illustration of UPF3b domains, showing the position of UPF2-binding region (blue) and EBM (green). Deletions and point mutations are depicted.

(legend continued on next page)

domain (U1BD) in the C terminus, whereas the binding to UPF3b is mediated by the third MIF4G domain (Figure 4D). We generated several point mutants in UPF2 to explore the function of these interactions during endocleavage. The mutation of glutamate at position 858 (E858R) has been previously reported to abolish the UPF2-UPF3b interaction in vitro and in cultured cells (Kadlec et al., 2004; Kashima et al., 2006). Notably, in our immunoprecipitation (IP) experiments, the E858R mutant was still able to co-IP UPF3b, albeit with reduced efficiency (Figure 4E). Therefore, we mutated more residues of UPF2 at the interface involved in UPF3b binding (Kadlec et al., 2004) and identified two variants of UPF2 (847-858^{Mut} and 847-867^{Mut}) for which the co-IP of UPF3b was reduced to background levels (Figure 4E). Importantly, both UPF2 mutants retained full interaction with UPF1 and we used UPF2 847-858^{Mut} for further analysis. We also confirmed that the previously characterized 1173^{Mut} construct (Clerici et al., 2009), in which the U1BD was mutated, did not bind UPF1 but still interacted with UPF3b in co-IPs (Figure 4E). In the UPF2 complementation assay, we were able to fully rescue NMD and endocleavage of the TPI-SMG5 mRNA with the UPF2 WT, whereas the 847-858^{Mut} construct only partially complemented the knockdown (Figures 4F and 4G). Strikingly, UPF2 1173^{Mut} was not able to complement the UPF2 depletion and no 3′ fragments were detected. This demonstrates that the UPF2-UPF1 interaction is absolutely required, whereas the interaction of UPF2 with UPF3b partially contributes to the endocleavage and degradation of mRNAs with long 3′ UTR.

EJCs Enhance Endocleavage of Long 3′ UTR mRNAs

In comparison to EJC-activated NMD substrates, only a small quantity of 3′ fragments was generated from mRNAs with long 3′ UTRs. To directly compare the efficiencies of endocleavage upstream of EJCs or long 3′ UTRs, we introduced a PTC at position 160 in the TPI reporter with a long 3′ UTR, which contains a single intron (Δi1-5) or lacks all the introns (Δi1-6; Figure 5A). Both reporters contain a long 3′ UTR, but the intron-containing TPI mRNA (Δi1-5) has an additional EJC downstream of the PTC, i.e., in the beginning of the 3′ UTR. In control conditions, we observed elevated levels of the EJC-containing mRNA (Δi(1-5); Figure 5B) comparable to our previous results (Figure 1K), which are likely caused by the general expression-enhancing function of EJCs. Notably, in XRN1 knockdown conditions, very small amounts of endocleavage products were detected for the intronless compared to the EJC-containing mRNA (Figure 5B, lanes 3 and 4), although the Δi1-6 construct produces detectable amounts of 3′ fragments when examined in isolation (Figure 2D). This indicates that the presence of an EJC activates NMD and thereby endocleavage to a larger extent than a long 3′ UTR alone. Markedly, the knockdown of UPF2 did not impede the endocleavage of the EJC-containing NMD substrate with a long 3′ UTR, which is in good agreement with our finding that UPF2 is required only for a subset of EJC-dependent endocleavage events (Figure 4A).

Tethering of NMD-Inducing Factors Promotes Endocleavage

The function of EJCs during NMD has been previously studied with tethering assays (Gehring et al., 2008). Owing to the strong stimulation of endocleavage by EJCs, we aimed to recapitulate this effect by tethering individual EJC components or NMD factors downstream of the termination codon. To this end, we modified the TPI-SMG5 reporter by introducing four MS2-binding sites between the termination codon and the SMG5 3′ UTR (Figure 5C). Tethering of the EJC core components Y14 and BTZ as MS2V5-tagged fusion proteins to the reporter resulted in a reproducible reduction of the reporter levels, confirming the functionality of our tethering system (Figures 5D and 5E). In XRN1-depleted cells, we detected low levels of 3′ fragments that were generated upstream of the long 3′ UTR. Strikingly, the basal endocleavage efficiency at the termination codon was strongly stimulated when Y14 or BTZ were tethered. This supports our earlier finding that EJCs stimulate endocleavage and demonstrates that this function of the EJC can be recapitulated by tethering of individual EJC components.

We used the same reporter construct to investigate whether endocleavage is also stimulated by tethered NMD factors, such as UPF2 and UPF3b. Tethering UPF2 WT decreased the levels of the reporter mRNA and strongly activated 3′ fragment generation, both comparable to the effects observed with EJC components. In contrast, UPF2 1173^{Mut}, which is deficient in UPF1 binding, was inactive in the tethering assay (Figures 5F, lane 8, and 5G), in agreement with the results of our UPF2 complementation assay (Figure 4F). Tethering of UPF2 847-858^{Mut}, which is unable to bind UPF3b, stabilized the reporter and even further decreased the low constitutive endocleavage to undetectable levels (Figures 5F and 5G). Hence, when this mutant of UPF2 was directly bound to the mRNA, it exerted a dominant-negative effect, which was not observed in the complementation assay (Figure 4F).

We next aimed to understand the role of UPF3b in SMG6-mediated endocleavage. Previous studies have shown that the depletion of UPF3b only moderately impairs degradation of different NMD substrates, potentially due to the functional substitution of UPF3b by UPF3a (Chan et al., 2009). In line with these previous studies, we observed mainly unaltered reporter mRNA levels upon a combined XRN1 and UPF3b knockdown using different TPI reporter mRNAs (Figures S4A–S4C). Interestingly, we detected increased endocleavage at several PTCs, which we did not observe with the reporter mRNA containing a long 3′ UTR. This unexpected finding prompted us to further characterize the effect of UPF3b on endocleavage in a tethering assay. To this end, we generated a variant of UPF3b with a K52D/RR56-57EE triple mutation (UPF3b 52/56^{Mut}; Figure 5H) in the region

(I) FLAG-UPF3b WT or mutant proteins were immunoprecipitated from RNase-A-treated HeLa cell lysates. Coprecipitated endogenous UPF2 was detected by immunoblotting.
(J) Northern blot analysis as in (D).
(K) Quantification of the mRNA levels in (J) as in (E).
See also Figure S4.

required for UPF2 binding (Kadlec et al., 2004). This UPF3b 52/56^{Mut} did not co-IP UPF2, whereas an UPF3b deletion mutant (Δ421-434), lacking the EBM and therefore deficient in EJC binding (Gehring et al., 2003), still interacted with UPF2 (Figure 5I). In tethering experiments, UPF3b WT and also UPF3b 52/56^{Mut} strongly activated endocleavage of the reporter mRNA, confirming the previous observation that the interaction with UPF2 is not required for UPF3b to activate degradation (Figures 5J and 5K; Gehring et al., 2003). However, we observed an upregulation of reporter mRNA levels and reduced endocleavage when the mutant of UPF3b lacking the EBM (UPF3b Δ421-434) was tethered. These results confirm that EJCs are critical for endocleavage and support the notion that the UPF3b-UPF2 interaction does not bridge the surveillance complex to a downstream EJC during NMD (Gehring et al., 2003).

3′ Fragments Are Polyadenylated and 5′ Phosphorylated

Previous studies demonstrated that the 3′ fragments generated by SMG6 are polyadenylated (Eberle et al., 2009; Gatfield and Izaurralde, 2004). The similarity of the SMG6 PIN domain with FLAP family proteins (Glavan et al., 2006) also suggests that SMG6 leaves a 5′ monophosphate at the 3′ fragment during endocleavage. Hence, it is likely that the 3′ fragments generated at PTCs in an EJC-dependent manner and 3′ fragments generated upstream of long 3′ UTRs have identical termini. To verify this assumption, we first performed a poly(A)-enrichment step with RNA samples obtained from XRN1-depleted cells transfected with TPI-Δi(1-5)-PTC160 or TPI-5. The 3′ fragments of both samples were efficiently precipitated, whereas rRNA was retained in the supernatant, showing that the 3′ fragments contain an accessible poly(A) tail (Figure 6A). To confirm the presence of the 5′ monophosphate, we performed an in vitro digest with recombinant yeast XRN-1 (rXRN-1), which only degrades 5′ phosphorylated RNAs in a highly processive manner. As expected, the treatment with rXRN-1 degraded the 3′ fragments of different reporter mRNAs, whereas the control and reporter mRNA itself remained intact (Figures 6B and 6C).

Mapping the Sites of Endocleavage

The 5′ monophosphate on the 3′ fragments makes them amenable to linker ligation at the 5′ end, and the poly(A) tail enables oligo(dT)-primed reverse transcription. On this basis, we developed a method to identify the cleavage sites of 3′ fragments by high-throughput sequencing (Figure 6D). Notably, the size of amplified 3′ fragments we obtained after the final PCR and prior to high-throughput sequencing corresponded to the length of 3′ fragments of different PTC-containing mRNAs (Figure 6E). We used the same strategy to amplify 3′ fragments of endogenous SMG5 mRNA.

Sequencing results had increased resolution and sensitivity compared to northern blotting and therefore allowed precise mapping of endocleavage sites. We sequenced the 3′ fragments of TPI mRNAs with different positions of PTCs in triplicates (Table S1). The reproducibility of the single replicates is shown for TPI-PTC189 (Figure S5A). For initial analysis, we assessed the density of reads in 5 nt windows of TPI mRNAs, such that the density in TPI mutants was normalized by subtraction of the TPI-WT, which we considered as background (Figure S5B). The normalized density of sequencing reads was highest in the region up to 100 nt downstream of the endocleavage positions in all TPI mutants (Figure S5B). Moreover, a nucleotide-resolution analysis of the PTC-proximal region (±100 nt around the PTC) showed that endocleavage events were rare upstream of the PTC but instead mapped to positions immediately downstream of the PTC (Figure S5C). We observed a mild preference for Gs in the area surrounding the cleavage sites (Figure S5D). Upon pooling the normalized reads from all PTC samples, we noticed two major cleavage sites around 0 and 24 nt downstream of the PTC (Figure 6F). Notably, high-throughput sequencing of 3′ fragments that were generated from the endogenous SMG5 mRNA showed a similar pattern with two very distinct peaks with maxima at 3 and 38 nt downstream of the stop codon (Figures 6G and S5E). In summary, single-nucleotide resolution analysis demonstrates that endocleavage occurs at two sites, one overlapping with the stop codon and another in the region of 20–40 nt downstream of the PTC, both in the TPI reporter and the endogenous SMG5 mRNAs.

DISCUSSION

Several mRNA-degradation pathways are involved in the removal of NMD substrates (Loh et al., 2013). These include nonsense-specific decay mechanisms, such as SMG6-mediated endocleavage, but also the recruitment of the general mRNA decay machinery. Despite several years of intensive work, the determinants and factors that govern the activation of individual decay routes are not well understood. In this work, we present a systematic analysis of SMG6-dependent endocleavage using different classes of NMD substrates and define the molecular characteristics of this NMD-specific decay pathway.

Splicing and EJCs Stimulate Endocleavage

Our analysis of mRNAs with PTCs in different exons clearly indicates that endocleavage occurs in close proximity to the PTC, which is in line with the results of previous studies (Eberle et al., 2009; Gatfield and Izaurralde, 2004; Huntzinger et al., 2008). However, we also observe a hitherto unrecognized stimulation of endocleavage in the presence of a downstream EJC. Notably, endocleavage is also stimulated by tethering individual EJC components, which confirms the specific effect of EJCs on endocleavage activity. Whereas our findings explain the NMD-activating function of EJCs, they also raise the question of how a distant EJC is detected during the initial phases of NMD. Current models suggest that ribosome-associated UPF1 is phosphorylated by its kinase SMG1 in response to a physical interaction with the EJC. Whichever molecular mechanism is responsible for the initial contact of UPF1 and the EJC, it must be able to bridge a large distance on the mRNA. Our data indicate that an EJC as far as 600 nt downstream of a PTC can still efficiently stimulate endonucleolytic cleavage at the PTC, which is in line with the observation that EJCs can exert long-range effects during NMD (Neu-Yilik et al., 2001). It remains to be determined whether such large distances are overcome by spreading of EJCs to noncanonical upstream positions or by a mechanism involving UPF1-dependent mRNA scanning, threading, or

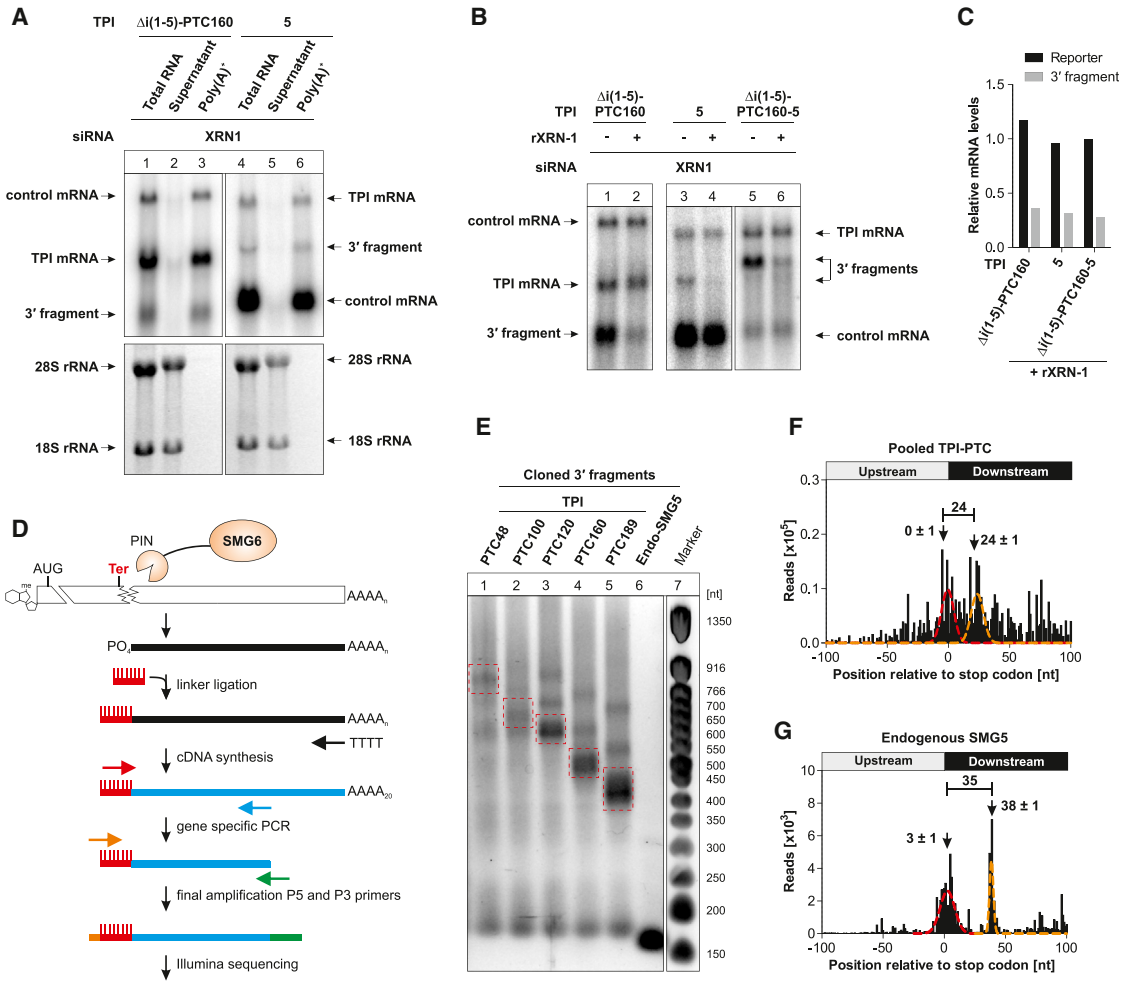


Figure 6. Mapping of 3′ Fragments Reveals Preference for Endocleavage at and Downstream of the Termination Codon

(A) Poly(A)$^+$ purification of total RNA extracted from XRN1 knockdown HeLa cells transfected with the indicated TPI reporters. Equal fractions of input total RNA, LiCl-precipitated unbound RNA, and poly(A)$^+$ RNA were analyzed by northern blotting. Ethidium-bromide-stained rRNA served as unbound nonpolyadenylated control (bottom).

(B and C) In vitro degradation assay of 5′ monophosphate RNA using recombinant yeast XRN-1 (rXRN-1). XRN1 knockdown cells were transfected with TPI reporter constructs; total RNA was isolated and treated with or without rXRN-1. Samples were analyzed by northern blotting (B) and mRNA levels quantified and normalized to untreated mRNA levels (C).

(D) Schematic overview of the 3′ fragment cloning strategy. The 5′ monophosphate containing mRNA endocleavage fragment is ligated to an RNA linker, followed by reverse transcription using oligo(dT) primer. Two cycles of PCR amplify specific 3′ fragments and introduce the sequences required for Illumina sequencing (P5 and P3 primers).

(E) Final PCR products before Illumina sequencing were separated on a 2.5% NuSieve GTG agarose Tris-borate-EDTA gel and stained with SYBR Gold. Red boxes indicate amplified 3′ fragments of the expected size for the respective PTC-containing reporter.

(F) Combined sequencing reads of PTC-proximal (±100 nt) 3′ fragments of all TPI reporter mRNAs are shown. For each reporter, the background (TPI-WT) was subtracted and the reads were normalized to the total read number of the respective reporter, pooled, and plotted against the position on the mRNA. The Gaussian distribution of the two major peaks has been calculated and is charted together with positions of their maxima and their distance.

(G) Reads of three replicates of endogenous SMG5 samples in the region ± 100 nt around the stop codon were combined, and major peaks are indicated as in (F). See also Figure S5 and Table S1.

looping to establish the physical interaction of UPF1 with the EJC. Interestingly, our data suggest that a hypothetical tag earmarks the position of the termination codon until endocleavage occurs in its vicinity. It will therefore be an important goal for future research to identify the tag and to explore the molecular mechanism that directs endocleavage to the position of translation termination.

In contrast to canonical NMD substrates that are degraded when at least one EJC is located downstream of the termination codon, long 3′ UTRs activate NMD in an EJC-independent manner. Nonetheless, our analysis of four reporter mRNAs with different 3′ UTR lengths provides clear evidence for endocleavage at the position of the termination codon. We show that this endocleavage is SMG6 dependent and can be stimulated by

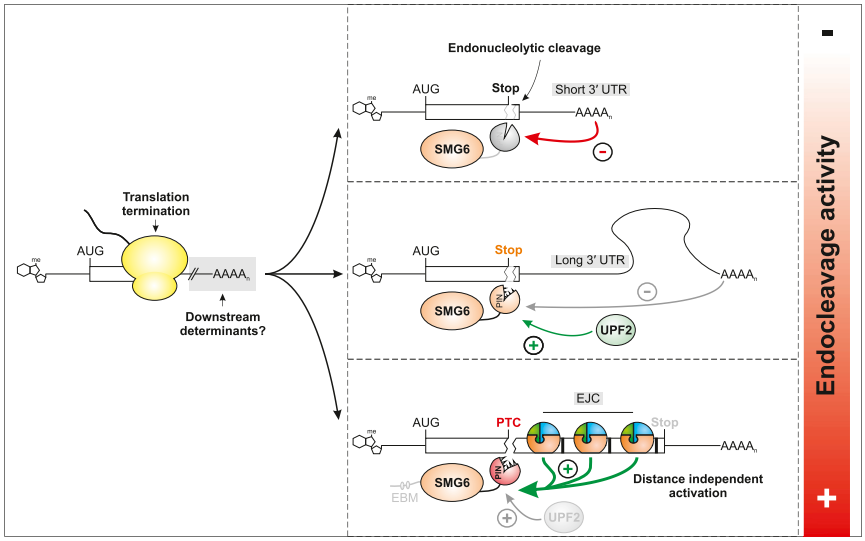


Figure 7. Model for SMG6-Mediated Endocleavage at Different Classes of Target mRNAs

Schematic representation of the influence of NMD factors and 3′ UTR composition on endocleavage efficiencies of various NMD substrates. For details, see Discussion.

the deposition of an EJC downstream of the termination codon. Hence, our data suggest that an endocleavage-competent NMD machinery is recruited upon translation termination upstream of long 3′ UTRs and EJCs.

Functions of SMG6 and UPF2 during Endocleavage

The PIN domain of SMG6 executes the endocleavage of the target mRNA during NMD (Eberle et al., 2009; Huntzinger et al., 2008). Because the EBMs of SMG6 have been reported to be essential during NMD (Kashima et al., 2010), we speculated that they directly or indirectly regulate the activity of the PIN domain. Whereas our complementation assays confirm the essential role of the PIN domain of SMG6 during endocleavage of EJC-dependent/independent NMD substrates, an EBM point mutant and EBM deletion mutants of SMG6, which are unable to interact with the EJC, exhibit wild-type endonuclease activity on the NMD substrates. Therefore, our results question the view that a direct interaction of SMG6 with the EJC is required during the endocleavage of NMD substrates.

UPF2 interacts with UPF1 and UPF3 in human cells and has been suggested to bridge EJCs and the NMD machinery during EJC-dependent NMD (Kashima et al., 2006). Hence, we hypothesized that the endocleavage-stimulating effect of the EJC is forwarded by UPF2. However, we find that UPF2 depletion inhibits endocleavage at long 3′ UTRs, whereas endocleavage at PTCs with downstream EJCs is mainly unaffected. UPF2 binding to UPF3b was also dispensable for endocleavage stimulated by tethered UPF3b. Although we observed a dominant-negative effect upon tethering of a UPF2 protein unable to interact with UPF3b, the complementation assay showed that this interaction is not absolutely necessary for endocleavage. Our findings challenge the current model of UPF2 functioning as the bridging factor between the surveillance complex and a downstream EJC but also argue against a direct link between SMG6 and the EJC. Importantly, we also provide evidence against redundant roles of UPF2 and the EBMs of SMG6 as bridging factors. Hence, the stimulation of endocleavage by EJCs either does not require a physical contact between EJC and the surveillance complex or involves as-yet-unknown proteins that mediate this contact.

Mapping of Endocleavage Sites

To compare the characteristics of endocleavage at different PTCs, we developed a strategy to isolate and amplify 3′ fragments. With this strategy, we observed that endocleavage occurs with the highest efficiency directly at the termination codon and closely downstream thereof. Importantly, we detect 3′ fragments generated at the termination codon of the endogenous SMG5 mRNA, which indicates that SMG6-dependent endocleavage efficiently targets endogenous long 3′ UTR NMD substrates. For both EJC-dependent (PTC reporters) and independent substrates (SMG5), we noticed two main sites of endocleavage: one overlapping with the stop codon and another at ~20–40 nt downstream. Compared to previous studies, this refines the sites of endocleavage by SMG6 and suggests that components of the NMD machinery are positioned closely downstream of the terminating ribosome. We propose that SMG6-mediated cleavage is thereby promoted at positions up- and downstream of the NMD complex. Furthermore, the presence of differently sized protein complexes, e.g., lacking or including UPF2, on different mRNA substrates may explain specific cleavage site preferences and hence distinct patterns of 3′ fragments. Additionally, our data indicate that the endocleavage step of NMD likely occurs after disassembly of the ribosome, because the region around the stop codon is not protected from cleavage. Indeed, endocleavage at the termination codon may represent a default mechanism of mRNA turnover. Such endocleavage is expected to occur at a very low frequency, which, besides the short length of the resulting 3′ fragments, provides an explanation why endocleavage at normal termination codons has not been described so far. Transcriptome-wide analyses of degradation intermediates will be required to test this hypothesis.

The previous work of several labs suggests that the composition of the messenger ribonucleoprotein downstream of the stop codon determines whether an mRNA is recognized as aberrant and is degraded by the NMD pathway. Based on our results, we suggest that the same signals that activate NMD are also used to determine the efficiency of endocleavage at the termination codon. A regular termination event upstream of a short 3′ UTR promotes efficient translation termination and ribosome recycling, thereby largely preventing degradation by SMG6-executed endocleavage (Figure 7, top panel). However, in the

presence of a long 3′ UTR, inefficient translation termination leads to the assembly of a UPF2-dependent surveillance complex and an increased frequency of endocleavage (Figure 7, middle panel). The highest frequency of endocleavage is observed in the presence of downstream EJCs, which strongly stimulate the endocleavage activity of SMG6 in a distance-independent manner. Notably, this effect does not require any physical bridging by UPF2 or the EBMs of SMG6 (Figure 7, bottom panel).

In conclusion, our results identify a single mechanism for the endonucleolytic decay of different classes of NMD substrates. Furthermore, we define the factors and sequence elements governing the generation of endocleavage products, thus providing methodological and biological insights into the molecular mechanisms of human NMD.

EXPERIMENTAL PROCEDURES

Plasmid Transfections

HeLa cells were split to 6-well plates the day after siRNA transfection and transfected by calcium phosphate precipitation with 0.5 μg of a GFP expression plasmid, 1.5 μg control plasmid (LacZ-4H), and 1.5 μg plasmid encoding for reporter mRNA. For long 3′ UTR reporters, 3 μg reporter and 0.75 μg control plasmid (β-globin) were transfected. For tethering or rescue assays, 1 μg of FLAG- or MS2V5-tagged expression plasmid was included in the transfection mix.

RNA Extraction and Northern Blotting

Total RNA was extracted with Isol-RNA Lysis Reagent (5PRIME) and analyzed by northern blotting as described (Gehring et al., 2009). Signals were quantified using a Typhoon FLA 7000 (GE Healthcare).

Immunoblot Analysis and Immunoprecipitation

SDS-polyacrylamide gel electrophoresis and immunoblot analysis was performed using protein samples derived from Isol-RNA Lysis Reagent extractions.

Magnetic M2 anti-FLAG beads (Sigma-Aldrich) were used to immunoprecipitate FLAG complexes from RNase A-treated (50 μg/ml) HeLa cell lysates in lysis buffer (50 mM Tris [pH 7.2], 150 mM NaCl, and 0.5% Triton X-100) supplemented with protease inhibitor (Sigma-Aldrich). Complexes were eluted with SDS-sample buffer, separated by SDS-PAGE, and analyzed by immunoblotting.

cDNA Library Preparation for High-Throughput Sequencing

Two hundred picomoles of RNA linker was ligated to 20 μg of total RNA with T4 RNA Ligase I. After RT-PCR with VNN oligo(dT)$_{20}$ primer, a gene-specific PCR using P5 and P3 Fusion primers (see Supplemental Experimental Procedures) was performed using Accuprime Taq Polymerase (Life Technologies). For the final PCR, P5 and P3 primers were used.

Analysis of High-Throughput Sequencing Data

Libraries of 3′ fragments were sequenced on the HiSeq 2500 machine from Illumina. For further details, see Supplemental Experimental Procedures.

ACCESSION NUMBERS

The ArrayExpress accession number for the 3′ fragment sequencing data reported in this paper is E-MTAB-2907. The input raw sequencing data and the detailed documentation are available from Bitbucket (https://bitbucket.org/nebo56/3-prime-fragment-sequencing).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.09.012.

AUTHOR CONTRIBUTIONS

N.H.G. and V.B. designed the study. V.B. performed the experiments. F.O. conceived and established the amplification of 3′ fragments. N.H. and J.U. carried out high-throughput sequencing. V.B., N.H., and J.U. analyzed and interpreted the resulting data, and V.B. and N.H.G. analyzed and interpreted all other data. V.B., J.U., and N.H.G. wrote the manuscript. All authors discussed the results and implications and commented on the manuscript at all stages.

ACKNOWLEDGMENTS

We thank Heidi Thelen and Juliane Hancke for technical assistance, the Leptin and Schnetz labs for sharing equipment, Gabriele Neu-Yilik for critical reading of the manuscript, and members of the N.H.G. lab for useful discussions. We also thank Chris Sibley for providing advice with cDNA library preparation and sequencing. We are grateful to Jens Lykke-Andersen for antibodies against UPF1 and UPF2. V.B. is supported by a fellowship from the IGSDHD. This research was funded by grants from the Fritz Thyssen Stiftung and the Deutsche Forschungsgemeinschaft (SFB635, B6) to N.H.G.

Received: May 1, 2014
Revised: August 5, 2014
Accepted: September 7, 2014
Published: October 9, 2014

REFERENCES

Bühler, M., Steiner, S., Mohn, F., Paillusson, A., and Mühlemann, O. (2006). EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3′ UTR length. Nat. Struct. Mol. Biol. *13*, 462–464.

Chakrabarti, S., Bonneau, F., Schüssler, S., Eppinger, E., and Conti, E. (2014). Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. Nucleic Acids Res. *42*, 9447–9460.

Chan, W.K., Bhalla, A.D., Le Hir, H., Nguyen, L.S., Huang, L., Gécz, J., and Wilkinson, M.F. (2009). A UPF3-mediated regulatory switch that maintains RNA surveillance. Nat. Struct. Mol. Biol. *16*, 747–753.

Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. Annu. Rev. Biochem. *76*, 51–74.

Clerici, M., Mourão, A., Gutsche, I., Gehring, N.H., Hentze, M.W., Kulozik, A., Kadlec, J., Sattler, M., and Cusack, S. (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. EMBO J. *28*, 2293–2306.

Clerici, M., Deniaud, A., Boehm, V., Gehring, N.H., Schaffitzel, C., and Cusack, S. (2014). Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2. Nucleic Acids Res. *42*, 2673–2686.

Culbertson, M.R. (1999). RNA surveillance. Unforeseen consequences for gene expression, inherited genetic disorders and cancer. Trends Genet. *15*, 74–80.

Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z., and Mühlemann, O. (2008). Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. PLoS Biol. *6*, e92.

Eberle, A.B., Lykke-Andersen, S., Mühlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat. Struct. Mol. Biol. *16*, 49–55.

Fatscher, T., Boehm, V., Weiche, B., and Gehring, N.H. (2014). The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay. RNA *20*, 1579–1592.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Mol. Cell *17*, 537–547.

Gatfield, D., and Izaurralde, E. (2004). Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. Nature *429*, 575–578.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., and Kulozik, A.E. (2003). Y14 and hUpf3b form an NMD-activating complex. Mol. Cell *11*, 939–949.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Mol. Cell *20*, 65–75.

Gehring, N.H., Hentze, M.W., and Kulozik, A.E. (2008). Tethering assays to investigate nonsense-mediated mRNA decay activating proteins. Methods Enzymol. *448*, 467–482.

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009). Disassembly of exon junction complexes by PYM. Cell *137*, 536–548.

Glavan, F., Behm-Ansmant, I., Izaurralde, E., and Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. EMBO J. *25*, 5117–5125.

Holbrook, J.A., Neu-Yilik, G., Hentze, M.W., and Kulozik, A.E. (2004). Nonsense-mediated decay approaches the clinic. Nat. Genet. *36*, 801–808.

Huntzinger, E., Kashima, I., Fauser, M., Saulière, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA *14*, 2609–2617.

Kadlec, J., Izaurralde, E., and Cusack, S. (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. Nat. Struct. Mol. Biol. *11*, 330–337.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev. *20*, 355–367.

Kashima, I., Jonas, S., Jayachandran, U., Buchwald, G., Conti, E., Lupas, A.N., and Izaurralde, E. (2010). SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. Genes Dev. *24*, 2440–2450.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nat. Rev. Mol. Cell Biol. *13*, 700–712.

Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. Genes Dev. *27*, 2125–2138.

Mühlemann, O., and Lykke-Andersen, J. (2010). How and where are nonsense mRNAs degraded in mammalian cells? RNA Biol. *7*, 28–32.

Neu-Yilik, G., Gehring, N.H., Thermann, R., Frede, U., Hentze, M.W., and Kulozik, A.E. (2001). Splicing and 3′ end formation in the definition of nonsense-mediated decay-competent human beta-globin mRNPs. EMBO J. *20*, 532–540.

Nicholson, P., Yepiskoposyan, H., Metze, S., Zamudio Orozco, R., Kleinschmidt, N., and Mühlemann, O. (2010). Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cell. Mol. Life Sci. *67*, 677–700.

Nicholson, P., Josi, C., Kurosawa, H., Yamashita, A., and Mühlemann, O. (2014). A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. Nucleic Acids Res. *42*, 9217–9235.

Nott, A., Le Hir, H., and Moore, M.J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. Genes Dev. *18*, 210–222.

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., and Ohno, S. (2012). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. Nucleic Acids Res. *40*, 1251–1266.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol. *6*, e111.

Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y., and Akimitsu, N. (2012). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol. *9*, 1370–1379.

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. EMBO J. *17*, 3484–3494.

Supplemental Information

3′ UTR length and mRNP composition determine endocleavage efficiencies at termination codons

Volker Boehm, Nejc Haberman, Franziska Ottens, Jernej Ule, Niels H. Gehring

Supplemental Information includes Supplemental Data (5 figures and 1 table) and Supplemental Experimental Procedures

Supplemental Data

Figure S1

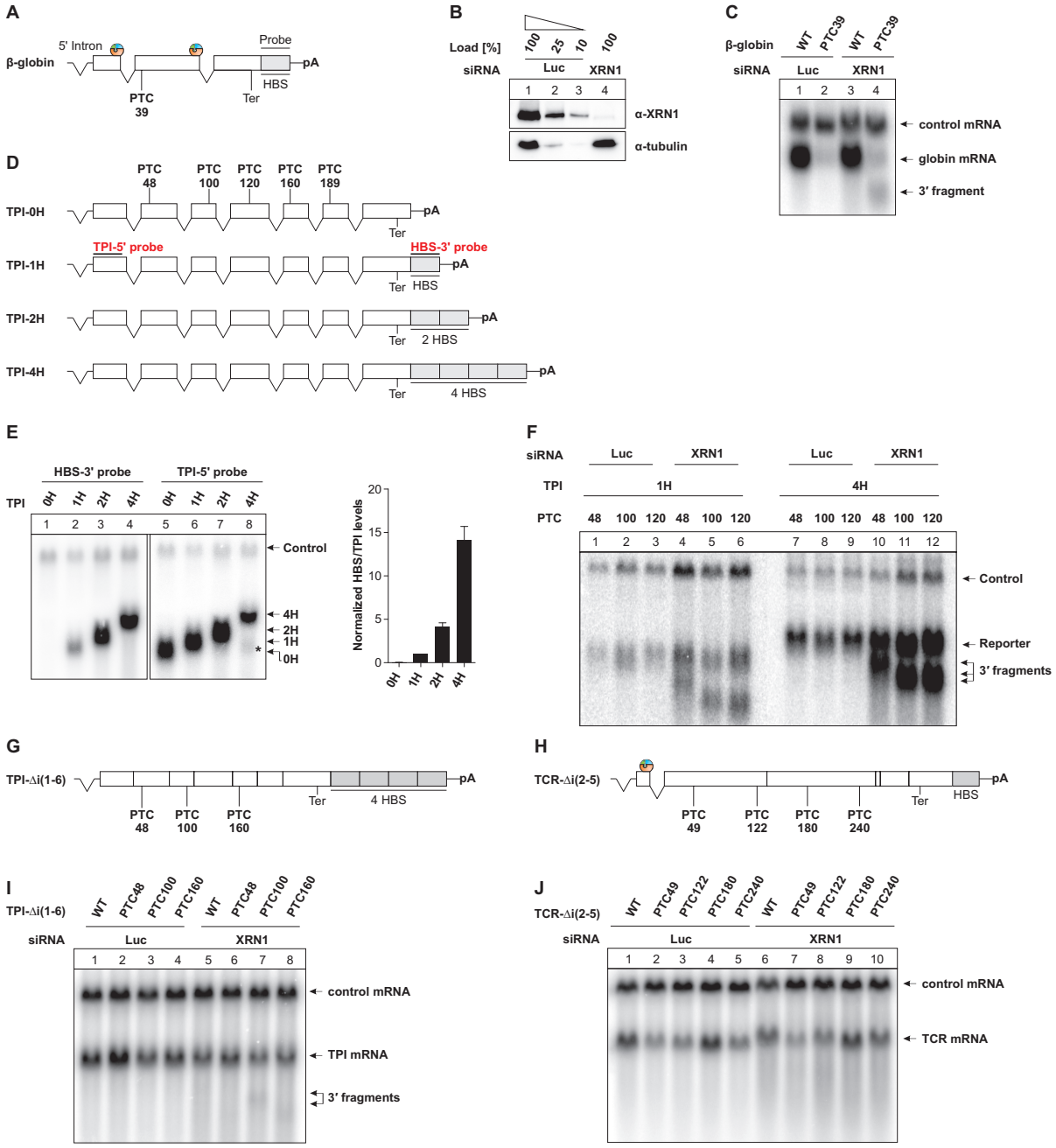


Figure S1, Related to Figure 1. XRN1 knockdown stabilizes 3′ fragments of PTC- and EJC-containing mRNA.

(**A**) Scheme of the β-globin reporter mRNA, showing the position of normal (Ter) stop codon and PTC, as well as canonical EJC deposition sites. The vector-derived 5′ UTR intron and poly(A) signal are illustrated. Hybridization position of the northern probe is indicated. (**B**) Western blot analysis of protein samples obtained from control (Luciferase) or XRN1 knockdown in HeLa cells. Tubulin serves as a loading control. (**C**) Northern blot analysis of total RNA derived from HeLa cells transfected with the indicated siRNAs and plasmids expressing β-globin WT or PTC reporter mRNA. An elongated β-globin WT mRNA (WT300+e3) serves as transfection control.

(**D**) Representation of TPI reporter mRNA without β-globin binding site (HBS; 0H) or multiple copies (up to 4H) fused to the 3′ UTR. Position of 5′ and 3′ northern blot probes are indicated. (**E-F**) HeLa cells were transfected with the indicated TPI reporters and RNA was analyzed by northern blotting using the heterologous probe targeting the HBS (**E-F**) or the TPI 5′ probe (**E**). Three independent experiments were quantified in (**E**) and the mean ± SD is plotted. The asterisk marks the position of endogenous TPI. TPI-5 (**E**) and LacZ-4H (**F**) served as transfection control. For (**F**), the indicated siRNAs were transfected prior to plasmid transfections. (**H-I**) Schematic overview of TPI and TCR reporter mRNAs lacking introns 1-6 and 2-5, respectively. (**J-K**) Northern blot analysis of RNA obtained from HeLa cells transfected with the indicated siRNA and plasmids expressing TPI (**J**) or TCR (**K**) reporter mRNA. LacZ-4H served as transfection control.

Figure S2

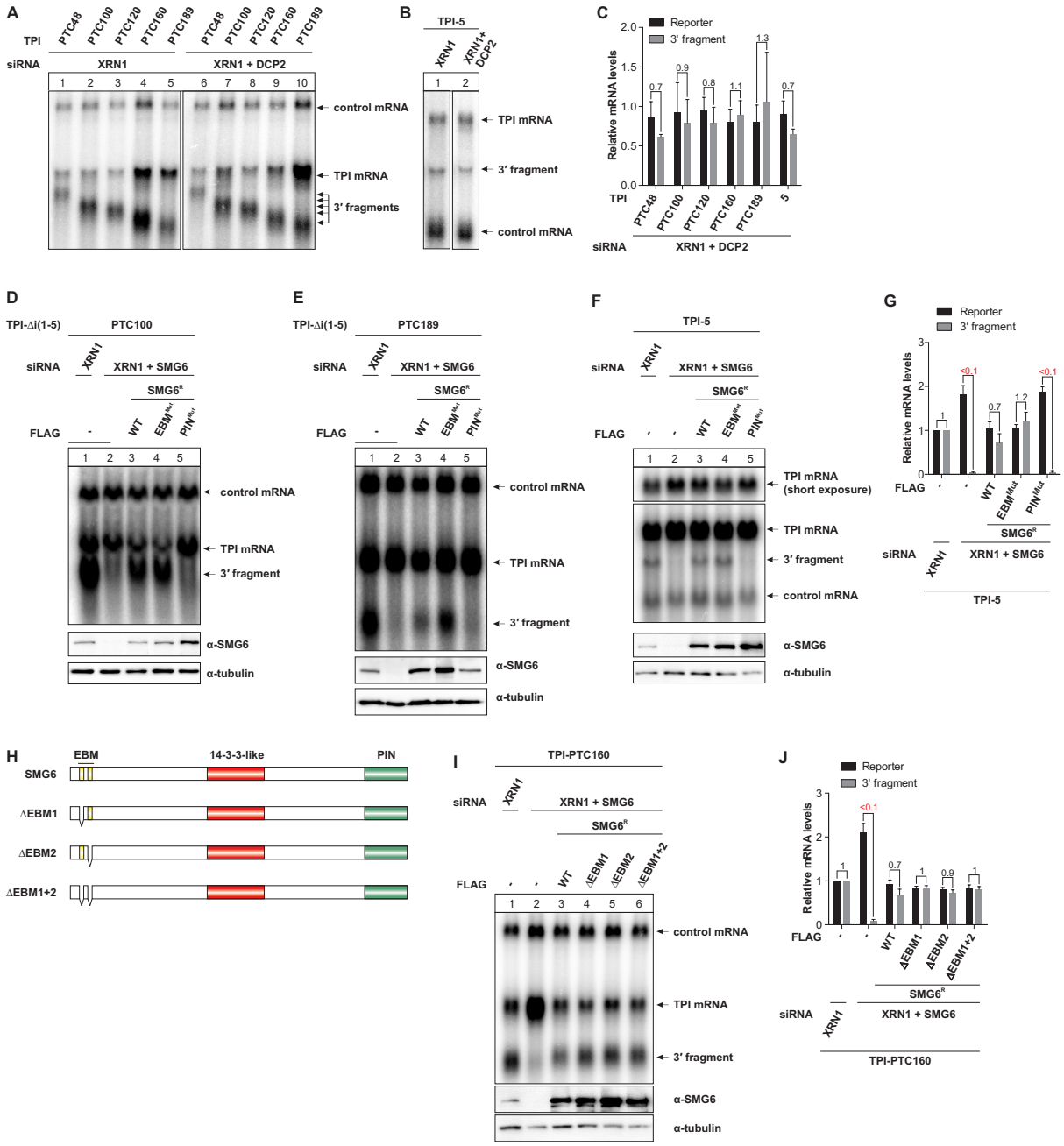


Figure S2, Related to Figure 3. Complementation assays for different PTC containing TPI reporter mRNA and SMG6 constructs.

(**A-C**) Transfection with the indicated siRNAs and reporter plasmids was performed in HeLa cells and total RNA was analyzed by northern blotting. mRNA levels (**A-B**) of XRN1 + DCP2 knockdown samples were quantified (mean ± SD; n = 3) and normalized to the respective XRN1 knockdown lanes (**C**). The ratio of 3′ fragment to reporter mRNA levels is indicated above the bars. LacZ-4H expressing plasmid was co-transfected as control mRNA. (**D-G, I-J**) Complementation assay of SMG6-depleted cells using TPI-Δi(1-5)-PTC100 (**D**), TPI-Δi(1-5)-PTC189 (**E**), TPI-5 (**F-G**) or TPI-PTC160 (**I-J**) as mRNA reporter. Knockdown and rescue with siRNA-resistant SMG6 constructs is shown by western blots, using tubulin as loading control (**D-F**, **I** lower panels). Northern blotting was performed of RNA extracted from the same cells used for western blot analysis (**D-F**, **I** upper panel). Mean values of reporter and

3′ fragment signal ± standard deviation of three independent experiments were quantified and normalized to the XRN1 control knockdown (**G** and **J**). The ratio of 3′ fragment to reporter mRNA levels is indicated above the bars. (**H**) SMG6 protein architecture is depicted schematically. EBM deletion mutants are indicated.

Figure S3

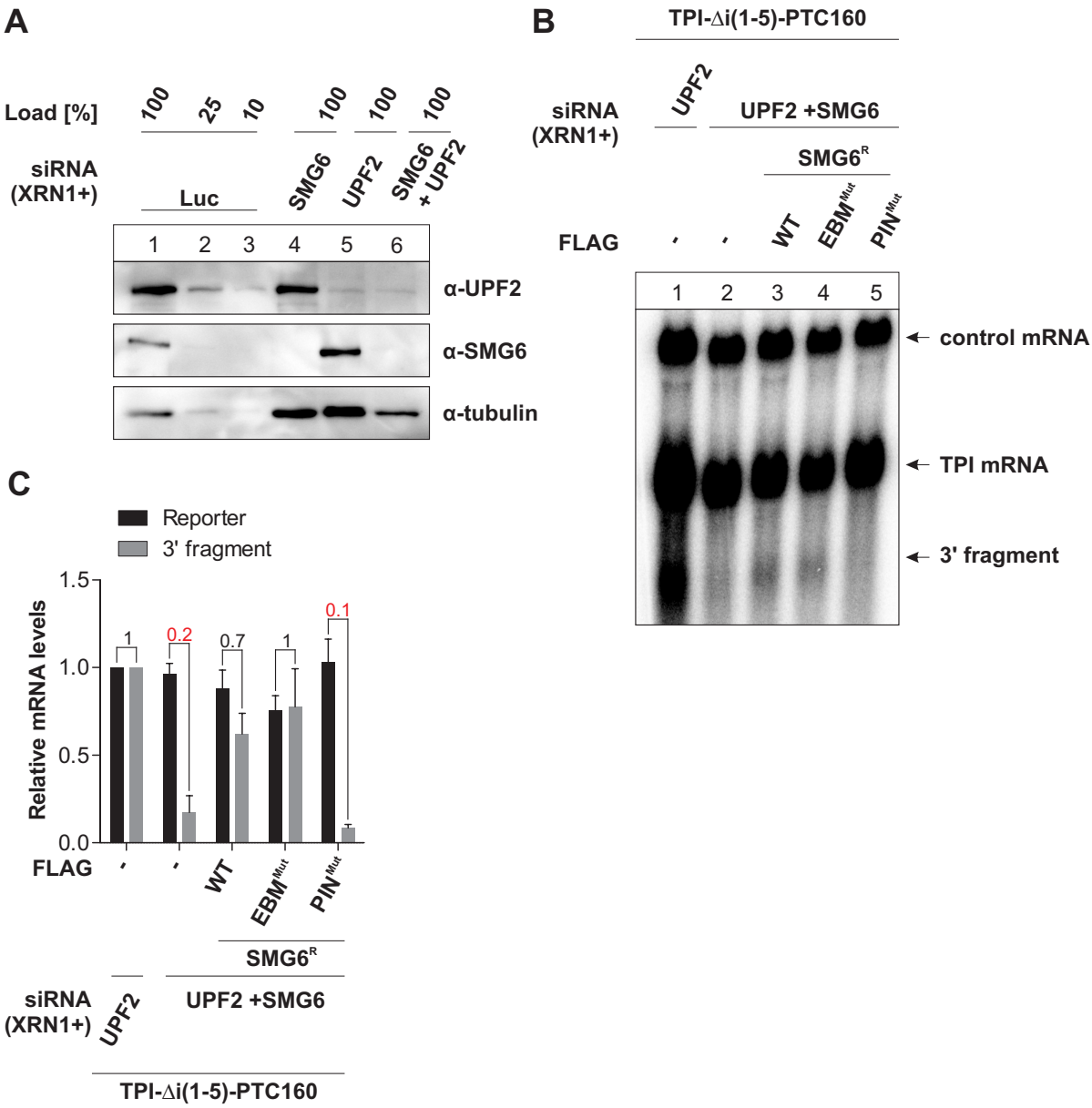


Figure S3, Related to Figure 4. UPF2 and SMG6 do not act as redundant bridging factors.

(**A**) Knockdown efficiencies in HeLa cells with the indicated mixtures of siRNA were analyzed by western blotting. Tubulin served as a loading control. (**B**-**C**) Complementation assay of XRN1-UPF2-SMG6-depleted cells using TPI-Δi(1-5)-PTC160 as mRNA reporter. Plotted in **C** is the mean ± SD, n = 3.

Figure S4

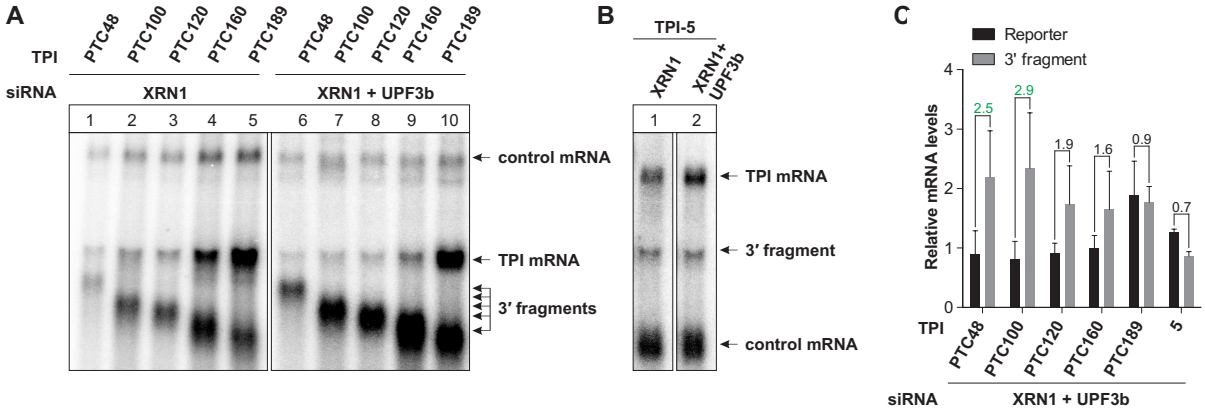


Figure S4, Related to Figure 5. Effect of UPF3b knockdown on endocleavage.

(**A-C**) Transfection with the indicated siRNAs and reporter plasmids was performed in HeLa cells and total RNA was analyzed by northern blotting. mRNA levels (**A-B**) of XRN1 + UPF3b knockdown samples were quantified (plotted is the mean ± SD, n = 3) and normalized to the respective XRN1 knockdown lanes (**C**). The ratio of 3′ fragment to reporter mRNA levels is indicated above the bars. LacZ-4H expressing plasmid was co-transfected as control mRNA.

Figure S5

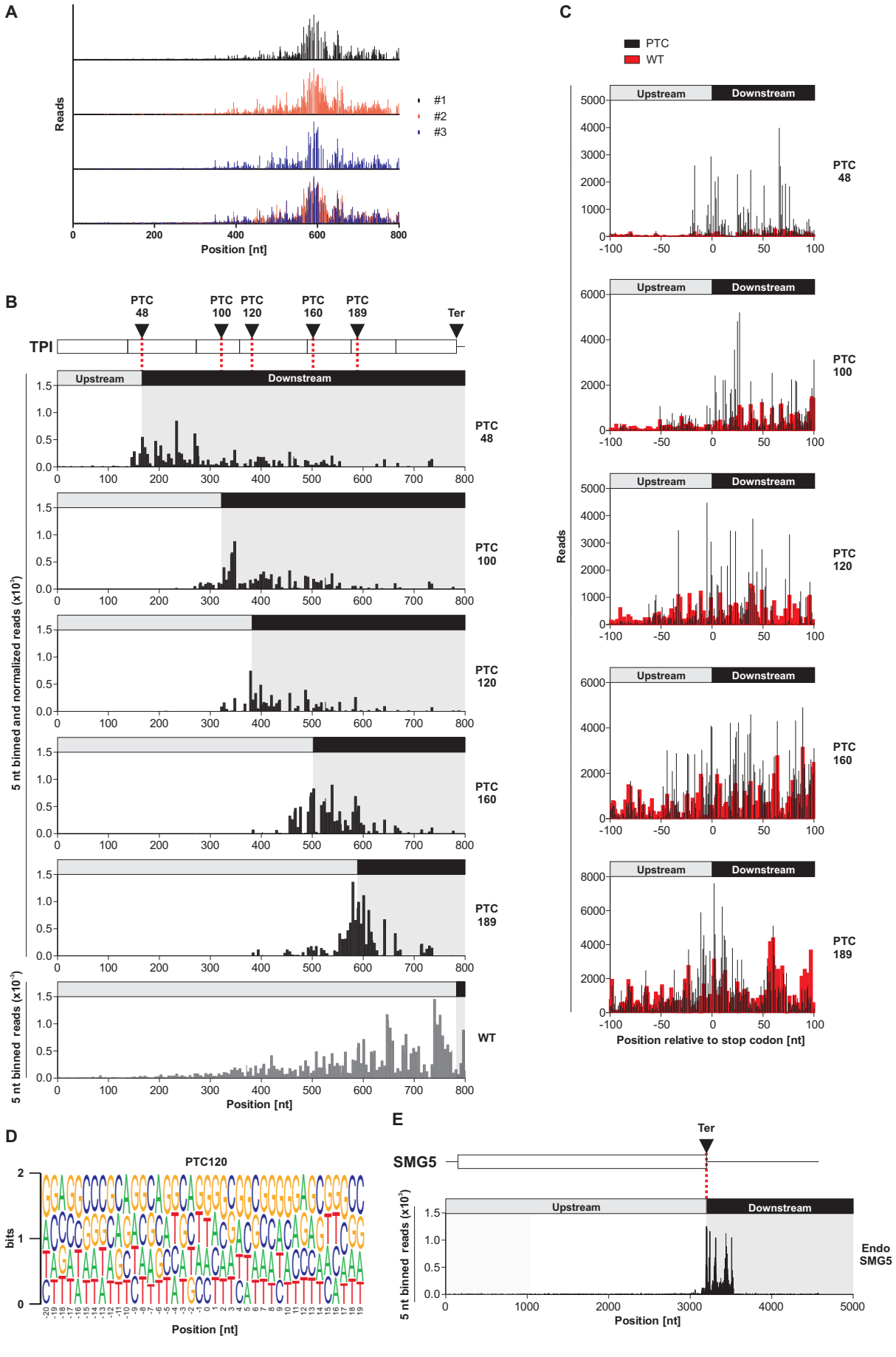

A
Reads
#1
#2
#3
0
200
400
600
800
Position [nt]
B
PTC 48
PTC 100
PTC 120
PTC 160
PTC 189
Ter
TPI
Upstream
Downstream
5 nt binned and normalized reads (x10⁻³)
5 nt binned reads (x10⁻³)
Position [nt]
C
PTC
WT
Reads
Position relative to stop codon [nt]
D
PTC120
bits
Position [nt]
E
SMG5
Endo SMG5
5 nt binned reads (x10⁻³)
Position [nt]

Figure S5, Related to Figure 6. Analysis of endocleavage patterns of amplified 3′ fragments.

(**A**) The sequencing results for the three individual replicates (black, orange and blue) of the TPI-PTC189 3′ fragments were plotted against the length of the mRNA, the overlay is shown as bottom panel. (**B**) Analysis of Illumina sequencing of amplified TPI reporter mRNA derived 3′ fragments. The reads from three replicates for each reporter were combined, binned to 5 nt, normalized to the TPI-WT reads (shown at the bottom) and plotted against the mRNA length. Regions upstream (gray) and downstream (black) of the PTC or stop codon are indicated. (**C**) Unprocessed reads for each reporter were plotted individually in a region 100 nt up- and downstream of the respective PTC. TPI-WT background signal is shown in red. (**D**) Weblogo of the motifs of 3′ fragments of TPI-PTC120. A region from the stop codon to 100 nt downstream was selected and for each position the TPI sequence ± 20 nt around the cleavage site was used as input for the Weblogo software (http://weblogo.threeplusone.com/). (**E**) Analysis of Illumina sequencing of endogenous SMG5 derived 3′ fragments. The reads from three replicates were combined, binned to 5 nt and plotted against the mRNA length.

Sample	Barcode	Total reads	0 times aligned	1 times aligned	>1 times aligned	Overall alignment
TPI-WT	GGTT	3640873	113407 (3.11 %)	3453194 (94.85 %)	74272 (2.04 %)	96.89 %
TPI-48	TTGT	3585582	178790 (4.99 %)	3299545 (92.02 %)	107247 (2.99 %)	95.01 %
TPI-100	CAAT	4314198	222428 (5.16 %)	4002566 (92.78 %)	89204 (2.07 %)	94.84 %
TPI-120	ACCT	4298651	247132 (5.75 %)	3959238 (92.10 %)	92281 (2.15 %)	94.25 %
TPI-160	GGCG	3802357	145296 (3.82 %)	3574438 (94.01 %)	82623 (2.17 %)	96.18 %
TPI-189	CCGG	4279615	148727 (3.48 %)	4064336 (94.97 %)	66552 (1.56 %)	96.52 %
Endo SMG5	TTAG	5772261	136730 (2.37 %)	5593111 (96.90 %)	42420 (0.73 %)	97.63 %
TPI-WT	AATG	4558091	211047 (4.63 %)	4259623 (93.45 %)	87421 (1.92 %)	95.37 %
TPI-48	TGGC	3858131	206103 (5.34 %)	3556974 (92.19 %)	95054 (2.46 %)	94.66 %
TPI-100	GGTC	4164955	219168 (5.26 %)	3864372 (92.78 %)	81415 (1.95 %)	94.74 %
TPI-120	AACC	4493269	262049 (5.83 %)	4135441 (92.04 %)	95779 (2.13 %)	94.17 %
TPI-160	CCAC	4881517	213287 (4.37 %)	4578675 (93.80 %)	89555 (1.83 %)	95.63 %
TPI-189	CGGA	4502097	137642 (3.06 %)	4301835 (95.55 %)	62620 (1.39 %)	96.94 %
Endo SMG5	GGCA	6279656	158329 (2.52 %)	6074804 (96.74 %)	46523 (0.74 %)	97.48 %
TPI-WT	AATA	4758532	205831 (4.33 %)	4475072 (94.04 %)	77629 (1.63 %)	95.67 %
TPI-48	TTAA	3864607	212178 (5.49 %)	3527323 (91.27 %)	125106 (3.24 %)	94.51 %
TPI-100	ATTT	3547887	180814 (5.10 %)	3289785 (92.73 %)	77288 (2.18 %)	94.90 %
TPI-120	CCTT	4515341	279297 (6.19 %)	4144777 (91.79 %)	91267 (2.02 %)	93.81 %
TPI-160	TATT	4370708	198061 (4.53 %)	4093747 (93.66 %)	78900 (1.81 %)	95.47 %
TPI-189	GCGT	4588353	168078 (3.66 %)	4361838 (95.06 %)	58437 (1.27 %)	96.34 %
Endo SMG5	AAGT	5621097	172804 (3.07 %)	5414339 (96.32 %)	33954 (0.60 %)	96.93 %

Table S1. Statistics of high-throughput sequencing.

Indicated are the mRNA samples from which the cloned 3′ fragments were derived, the barcode used in the P5 fusion primer, the total number of reads and mapping statistics.

Supplemental Experimental Procedures

Plasmids and cell culture

Plasmid constructs β-globin WT and PTC39, pCI-FLAG, pCI-Venus, WT300+e3, pCI-TPI, pCI-TCR-HBB, LacZ-HBB and expression vectors for UPF3b, BTZ and Y14 were described previously (Gehring et al., 2005; Gehring et al., 2009; Gehring et al., 2003; Steckelberg et al., 2012). Single or multiple (up to 4) copies of a 100 nt sequence of β-globin third exon and 3′ UTR, were cloned in cassettes downstream of the reporter sequence. Using the same cassette cloning strategy, HeLa cDNA-derived 3′ UTRs of SMG5 and UPF3b, GFP coding sequence or 4MS2 binding sites were introduced in the vectors. Intron-less reporter constructs were generated by PCR mutagenesis, either deleting introns (TCR) or inserting introns into TPI cDNA. All reporter mRNAs expressed from the pCI-neo contain a single intron in their 5′ UTR, which is encoded by the vector sequence. Point mutants of SMG6, UPF2 and UPF3b were generated by site-directed mutagenesis, cloned in the designated vector and verified by sequencing. siRNA insensitive UPF2 constructs were described previously (Clerici et al., 2013) and SMG6 was mutated accordingly (targeting sequence 5′-GTCACAGTGCTGAAGT -3′ replaced by 5′- GTGACCGTCCTCAAA -3′).

siRNA transfections

$5x10^5$ HeLa cells were grown over night in 6 cm plates and transiently transfected with 300 pmol siRNA for single or 600 pmol total siRNA for double knockdowns using Lipofectamine RNAiMAX (Life Technologies). 24 h post transfection the cells were split 1:2 in 10 cm plates and the day after transfected again with 600 pmol (single knockdown) or 1200 pmol (double knockdown) siRNA. For triple knockdown, 400 pmol of the single siRNAs were used (1200 pmol in total). The following siRNA target sequences were used for luciferase 5′-CGTACGCGGAATACTTCGA-3′, for XRN1 5′-AGATGAACTTACCGTAGAA-3′, for SMG6 5′-GGGTCACAGTGCTGAAGTA-3′ and for UPF2 5′-CGTTGTGGATGGAGTGTTA-3′, for DCP2 5′-GGACTGGCTTTCTCGAAGA-3′, for UPF3b an equal mix of 5′-GAGCATACATCAACTTTAA-3′ and 5′-GGAATATCCCGCTATAGTA-3′.

Expression and purification of recombinant proteins

N-terminally GST- and C-terminally FLAG-tagged SMG6 (1-207) wild type and mutant constructs were transformed in *E.coli* Rosetta 2 (DE3) pLysS (Novagene) cells. Bacteria were grown in LB medium at 37 °C until OD_{600} = 0.6 – 0.8. Protein expression was induced with 0.2 mM IPTG and the cells were grown for 20 h at 20 °C. The cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT) supplemented with 0.2 mg/ml lysozyme, 5 µg/ml DNase I, 0.5 mM EDTA, 0.5 mM Pefabloc, 1x Protease Inhibitor (Sigma-Aldrich) and proteins were purified via affinity chromatography using Glutathione HiCap columns (Qiagen). Elution was performed with lysis buffer supplemented with 25 mM L-glutathione and proteins were buffer exchanged in 25 mM Tris-HCl pH 8.0, 250 mM NaCl, 1 mM DTT.

In vitro immunoprecipitation analysis

7.5 µg of eIF4A3, untagged or FLAG-tagged BTZ and 10 µg of MAGOH/Y14 dimer were incubated in rEJC buffer (20 mM HEPES pH 7.5, 125 mM NaCl, 2 mM $CaCl_2$, 2 mM NaAc, 1 mM $MgCl_2$, 10 % glycerol, 1 mM DTT) supplemented with 10 µM poly(U)$_{15}$ RNA and 4 mM AMPPNP. Pre-assembly of the EJC was performed for 6 h at 4 °C. 12.5 µg FLAG-tagged SMG6 WT or mutants were added to the EJC and incubated over night at 4 °C. 20 µl of pre-washed anti-FLAG beads (50 % slurry; M2 magnetic beads; Sigma-Aldrich) were added and incubated for 2 h at 4 °C in rEJC buffer containing 0.1 % NP-40 (wash buffer). Beads were washed three times with 600 µl wash buffer and co-precipitated proteins were eluted with 1x SDS loading buffer. 10 % of the protein mix was used as input control, all samples were separated on 12 % SDS PAGE gels and stained with Coomassie Brilliant Blue.

Antibodies

The antibodies against tubulin (T6074) and FLAG (F7425) were from Sigma, the antibody against V5 (18870) was from QED Bioscience, the antibodies against GFP (ab290) and SMG6 (ab87539) were from Abcam, the antibody against XRN1 (A300-443A) was from Bethyl and the antibodies against UPF1 and UPF2 were kindly provided by Jens Lykke-Andersen. UPF3B antiserum was raised in rabbits by Eurogentech against a C-terminal fragment of UPF3B (300-483) and affinity purified.

Poly(A)$^+$ and rXRN-1 assays

250 µg of total RNA samples extracted from HeLa cells transfected with the indicated reporter and control plasmids were used for isolation of poly(A)$^+$ RNA with the magnetic mRNA isolation kit (New England Biolabs). Unbound RNA was LiCl precipitated and fractions of input, supernatant and elution were analyzed by northern blotting. For rXRN-1 assay, 4 µg of total RNA was incubated with or without 1 unit of rXRN-1 (New England Biolabs) for 2 h at 37 °C. RNA was precipitated with LiCl and analyzed by northern blotting.

Analysis of high-throughput sequencing data

High-throughput sequencing of 3′ fragment libraries was performed in two lanes of an 8-lane flow cell on the HiSeq 2500 machine from Illumina. Before mapping the reads, we removed random barcodes and adapter sequences. Mapping of reads against the appropriate transcript was performed by using Bowtie2.1 alignment software where we did not allow any mismatches. More than 90 % of reads mapped uniquely to the appropriate transcript. If multiple reads with the same random barcode mapped to the same starting position on the transcript, these were considered to be artefacts of variable PCR amplification and were collapsed to a single read. The position at the beginning of mapped reads was used to produce RNA maps. For normalized maps, we analyzed bins of 5 nucleotides, where the number of reads in TPI-WT (considered as background) was subtracted from the TPI mutants. The package of scripts performing the full analysis, along with the input raw sequencing data and the detailed documentation are available from Bitbucket (https://bitbucket.org/nebo56/3-prime-fragment-sequencing).

List of primers used for 3′ fragment cloning protocol.

RNA Linker VB	rGrCrUrGrArUrGrGrCrGrArUrGrArArUrGrArNrNrNrNrNrArArA
VNN oligo(dT)$_{20}$	TTTTTTTTTTTTTTTTTTTTVNN
P5 Fusion #1	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTGCTGATGGCGATG AATGA
P5 Fusion #2	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGTGCTGATGGCGATG AATGA
P5 Fusion #3	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCAATGCTGATGGCGATG AATGA
P5 Fusion #4	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTACCTGCTGATGGCGATG AATGA
P5 Fusion #5	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCGGCTGATGGCGATG AATGA
P5 Fusion #6	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGGGCTGATGGCGATG AATGA
P5 Fusion #7	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTTAGGCTGATGGCGATG AATGA
P5 Fusion #8	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAATGGCTGATGGCGATG AATGA
P5 Fusion #9	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGCGCTGATGGCGATG AATGA
P5 Fusion #10	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTCGCTGATGGCGATG AATGA
P5 Fusion #11	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCGCTGATGGCGATG AATGA
P5 Fusion #12	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCCACGCTGATGGCGATG AATGA
P5 Fusion #13	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGAGCTGATGGCGATG AATGA
P5 Fusion #14	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCAGCTGATGGCGATG AATGA
P5 Fusion #15	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAATAGCTGATGGCGATG AATGA
P5 Fusion #16	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTTAAGCTGATGGCGATG AATGA
P5 Fusion #17	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATTTGCTGATGGCGATG AATGA
P5 Fusion #18	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTGCTGATGGCGATG AATGA
P5 Fusion #19	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTATTGCTGATGGCGATG AATGA
P5 Fusion #20	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGTGCTGATGGCGATG AATGA
P5 Fusion #21	TCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGTGCTGATGGCGATG AATGA
P3-TPI primer	GATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTGGGGTGCTCGA GTCC
P3-SMG5 primer	GATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGATCCAAGAACC CATTCCAGT
P5 Solexa	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCT
P3 Solexa	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTC TTCCGA

3. Discussion

Several quality control mechanisms constantly monitor the mRNP integrity and architecture during each step of gene expression in order to maintain the fidelity of the transcript-encoded information. To assure the accurate production of full-length proteins, NMD identifies and degrades mRNPs with premature or unusual translation termination. The role of NMD is not only to protect the cell from the accumulation of potentially harmful truncated peptides, but also to regulate error-free mRNPs via certain NMD-activating signals. Despite more than a decade of intensive research, it is still not fully understood how these signals are to be interpreted, integrated in the NMD machinery and finally result in the degradation of the mRNP.

This cumulative work provides detailed new insights into (1) the mechanism of premature stop codon detection (Fatscher et al., 2014), (2) the deposition of EJCs during splicing (Steckelberg et al., 2012), (3) the interplay of NMD factors (Clerici et al., 2014) and (4) the requirements for initiation of the endonucleolytic degradation pathway (Boehm et al., 2014). Taken together, these studies provide a substantial improvement of our current NMD model and simultaneously identify critical points, which need to be investigated in the future.

In the following, the advancements and remaining uncertainties in the understanding of the NMD pathway are discussed, covering the events from translation termination and NMD assembly to mRNA degradation. Although not all details of the newly gained insights into the NMD pathways can be displayed in a clearly represented way, the following model is intended to give an updated view of the NMD mechanism (Figure 9).

3.1 The long 3′ UTR mRNP composition influences NMD activation

During each round of translation, the ribosome stops at a termination codon and requires a set of protein factors to release the peptide chain and disassemble the ribosome. Moreover, each time a stop codon enters the A site of the ribosome, downstream elements dictate whether the transcript is supposed to be degraded or kept for another round of translation. In mammals, EJCs and long 3′ UTRs are the two major NMD-inducing mRNP elements identified.

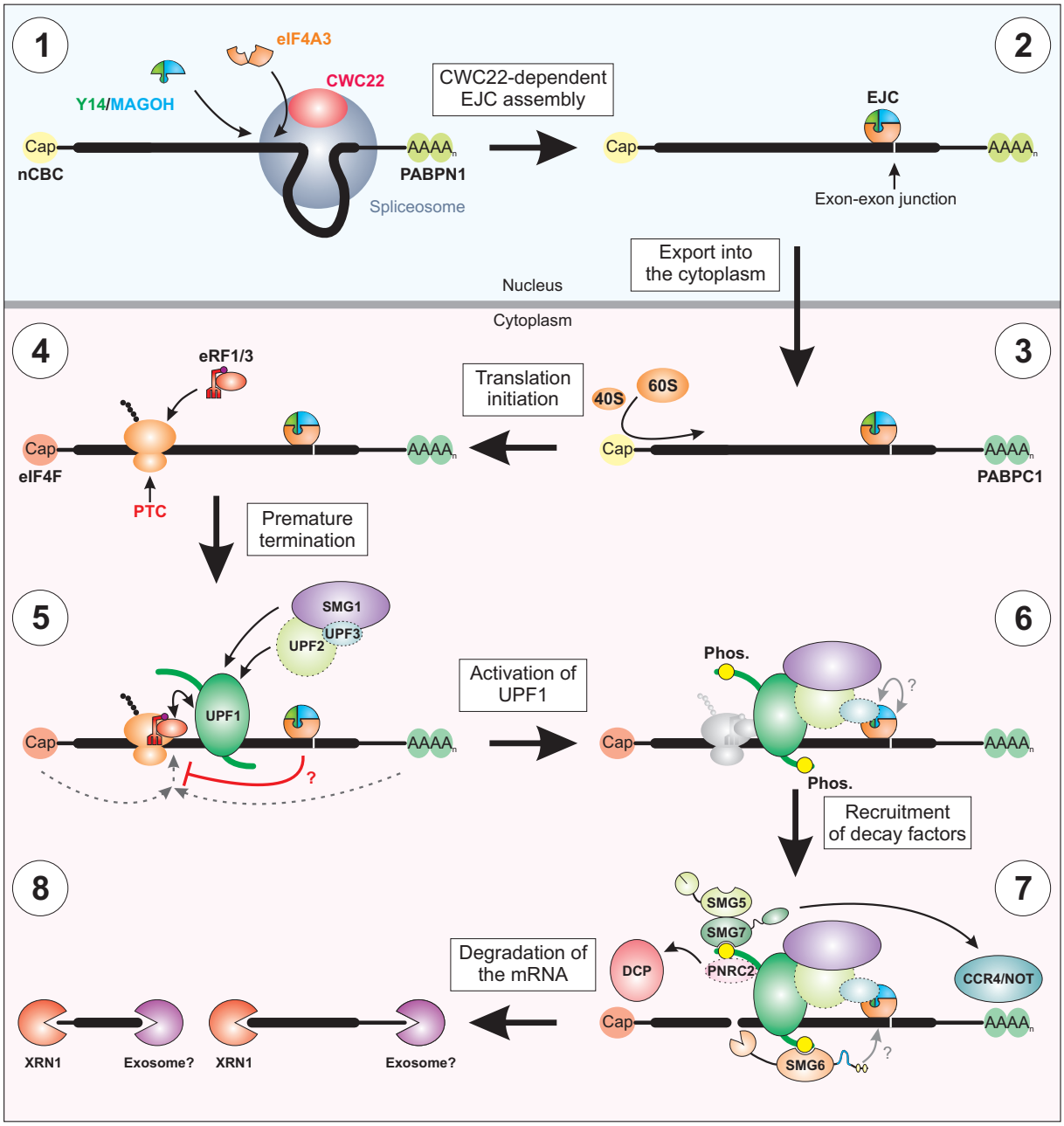


Figure 9: Improved model of mammalian NMD. The indicated mRNP elements and factors are depicted as in Figure 1 and Figure 8. For details, see Discussion.

When analyzed from a mechanistic point of view, NMD activation via long 3′ UTRs is only poorly understood. According to the faux 3′ UTR model, the competition between PABPC1 and UPF1 for eRF3 binding is the main determinant for NMD suppression or activation. In the presented work, it was confirmed that tethering of PABPC1 closely downstream of a stop codon inhibits the degradation-inducing effect of a long 3′ UTR (Fatscher et al., 2014). Contrary to the common model, the interaction of PABPC1 with eRF3 was not essential for the NMD suppression conferred by tethered PABPC1. Conversely, tethered eRF3 required the intact PABPC1-interacting PAM2 motif for the inhibition of reporter mRNA degradation. Interestingly, no induction of NMD was observed upon tethering of PABPC1-binding deficient eRF3 F76A

mutant. Using co-immunoprecipitation experiments, it was shown that the binding sites for PABPC1 (N-terminus) and UPF1 (GTPase domain) do not overlap, suggesting that the eRF3 F76A mutant is still able to interact with UPF1 (Ivanov et al., 2008). According to the prevalent model, this would lead to strong NMD activation, which was not observed. A more technical explanation for this phenomenon is that tethered eRF3 might not be able to enter the ribosome and therefore cannot properly initiate normal translation termination. This might be required, as the eRF3-UPF1 interaction alone might not be sufficient for NMD activation, which is a hypothesis that is currently difficult to implement in the common NMD models. To explore this more in depth, complementation assays in eRF3-depleted cells could be performed. A complication for this kind of approach is that two homologs of eRF3 exist in mammalian cells, eRF3a and eRF3b, both of which can fulfill the function to mediate translation termination (Jakobsen et al., 2001). For various cell lines and mouse tissues, readily detectable eRF3a protein levels were observed, whereas eRF3b was found to be expressed only in mouse brain and not in commonly used cell lines (including HeLa and HEK 293 cells). It was not investigated whether eRF3b is upregulated upon eRF3a depletion, however, functional studies showed that only siRNA-mediated silencing of eRF3a, but not eRF3b, had an effect on read-through efficiency (Chauvin et al., 2005). Therefore, eRF3a is the logical candidate for further investigations, although eRF3b should be inactivated as well to prevent any potential compensatory effects.

The interaction with the cytoplasmic cap-binding complex component eIF4G was necessary for PABPC1-mediated NMD suppression in tethering assays. This unexpected finding adds a new layer of complexity to the model of long 3′ UTR induced NMD and to normal translation termination in general. The eIF4F complex plays an important role in translation initiation and eIF4G represents the versatile scaffolding factor in this trimeric complex consisting of eIF4A, eIF4E and eIF4G (Hinnebusch, 2014). The eIF4G-PABPC1 interaction is required for the circularization of the mRNA, which positively stimulates translation (Sonenberg and Hinnebusch, 2009; Wells et al., 1998). It is conceivable that NMD-suppression requires specific interactions involving eIF4F, in order to stimulate the eRF3-regulated progression of translation termination. Alternatively, the close proximity of eIF4G to the terminating ribosome could promote efficient ribosome recycling by allowing reinitiation at the 5′ end of the mRNA. Supporting this hypothesis, it was recently shown that the presence of certain translation initiation factors was sufficient to promote reinitiation of 40S post-termination ribosomal

subunits (Skabkin et al., 2013). eIF4G interacts directly with multiple subunits of the eIF3 complex, which was one of the important factors for reinitiation, suggesting that eIF4G indirectly facilitates the recycling of ribosomal subunits (Villa et al., 2013). *In vitro* experiments revealed that under certain conditions, eIF3 together with eIF1 and eIF1A can promote ribosome recycling independently of other factors like ABCE1 (Pisarev et al., 2007; Pisarev et al., 2010). Moreover, studies in yeast showed that one peripheral component of the eIF3 complex, eIF3j, is involved in the dissociation of eRF3:GDP from the ribosome and thereby promotes translation termination (Beznoskova et al., 2013). Along this line, the NMD-suppressing effect of tethered eIF4G was decreased upon knockdown of eIF3 components (Joncourt et al., 2014). Consequently, the improved model for normal and NMD-inhibiting translation termination involves the interaction between eRF3 and PABPC1, which in turn recruits the eIF4F complex via eIF4G. This probably leads to the generation of a "figure of eight"-like structure, with the 5′ and 3′ ends of the mRNA, bound by eIF4F and PABPC1, recruited together to the ribosomal stop site (Figure 9, step 5). Multiple lines of evidence indicate that eIF3-mediated accelerated translation termination and/or ribosome recycling achieves fast and efficient clearance of stalled ribosomes. Whereas this cascade of events improves our understanding of normal translation termination, we still lack important insight into the mechanism by which UPF1 and subsequently the whole NMD pathway is activated upon PTC detection. One hypothesis is that UPF1 binding to eRF3 induces conformational changes in UPF1 or exposes binding sites for additional factors, such as SMG1. Furthermore, it will be important to dissect under which circumstances the eRF3-UPF1 interaction can occur and which of these binding events are NMD-activating. A caveat is the lack of specific point mutants of either eRF3 or UPF1, which do not interact with the other binding partner. All interaction data obtained so far used deletion mutants, which disrupt regions with multiple molecular roles and therefore cannot be used in functional assays.

It remains unclear for which molecular reason the presence of a long 3′ UTR result in aberrant translation termination. This is because calculations based on approximately 15000 human RefSeq mRNA 3′ UTRs estimated that the average 3′ UTR length is around 1300 nucleotides (Zhao et al., 2011). Since the insertion of the GFP coding sequence (719 nucleotides) is already enough to trigger NMD, it is surprising that the majority of endogenous mRNAs are not degraded via NMD (Boehm et al., 2014; Singh et al., 2008). It is therefore questionable whether long 3′ UTR NMD targets are truly degraded only because of the long distance between stop

codon and poly(A) tail. In a recent publication, several endogenous mRNAs containing 3′ UTRs of a length that is usually sufficient to activate NMD were investigated to assess their potential to be degraded via the NMD pathway. It was determined that many of these transcripts are insensitive to NMD and it was proposed that the A/U content is an important factor for the degradation-evading mechanism (Toma et al., 2015). The underlying principle is unclear, but it was discussed that different trans-acting factors are recruited when the nucleotide composition close to the stop codon is altered. Depending on the type of trans-acting factor, either NMD is stimulated or inhibited. As one potential candidate, it was shown that PABPC1 associates with high A/U content RNA, which could result in a functionally similar effect as the artificial tethering of PAPBC1 (Bollig et al., 2003; Sladic et al., 2004). Conversely, AU-rich regions could prevent or decrease the association of UPF1, since genome wide studies showed a preference of UPF1 for binding to GC-rich sites (Hurt et al., 2013). Most certainly, the involvement of many possible trans-acting factors results in a unique mRNP composition for any given mRNA, which requires intensive genome-wide analyses to uncover the central determinants.

3.2 EJC loading on the mRNP and the involvement in NMD

Contrary to long 3′ UTRs, which represent a heterogeneous group of NMD targets with potentially vastly different mRNP compositions, the EJC as NMD-inducing element is a well-defined mark on the mRNA. EJCs in general are important components of spliced mRNPs and regulate multiple gene expression processes. Early studies showed that single EJCs are deposited during splicing 20-24 nucleotides upstream of the exon-exon boundary in a sequence-independent manner (Le Hir et al., 2000). Recent iCLIP data showed that while this rule still holds true for most splicing events, not all canonical EJC sites are occupied and EJCs also have the potential to form higher molecular weight multi-EJCs (Sauliere et al., 2012; Singh et al., 2012). Despite detailed knowledge about the biochemical mechanism of how the EJC is attached tightly on the mRNA, the question remained how and by which protein the EJC is positioned there in the first place. In this work, the essential splicing factor CWC22 was identified as the EJC-loading factor (Figure 9, step 1) (Steckelberg et al., 2012). Specifically, mutations in eIF4a3, which abolish the interaction with CWC22, rendered eIF4A3 unable to assemble EJCs in splicing assays. The proposed role of CWC22 as the spliceosomal EJC assembly factor was supported by two independent publications from other labs (Alexandrov et al., 2012; Barbosa et al., 2012). Interestingly, the eIF4A3 interaction with CWC22 is mutually exclusive

with the binding of the Y14/MAGOH heterodimer, indicating that eIF4A3 loading on the mRNA occurs before stable formation of the EJC. Recent structural work provides additional insight and shows that CWC22 binding to eIF4A3 keeps the ATPase in an inactive state (Buchwald et al., 2013). This allows for the correct positioning of eIF4A3 until Y14/MAHOH lock the core EJC on the mRNA.

When ribosomes stop translation sufficiently upstream of an EJC, strong NMD activation is observed (Boehm et al., 2014). According to the prevalent model of NMD, the EJC induces or accelerates the formation of the NMD complex. This is supposedly achieved by the UPF3b-mediated recruitment of UPF2. UPF2 in turn can either modulate the helicase activity of UPF1 or promote the phosphorylation of UPF1 via SMG1 activation. While this model is in principle straightforward and could explain the strong induction of NMD, several lines of evidence speak against this simple recruitment role. First, tethering of UPF3b requires the interaction with the EJC to induce the degradation of the reporter mRNA (Boehm et al., 2014; Gehring et al., 2003). If the NMD-activating effect of the EJC is channeled only via UPF3b and subsequently UPF2, we should observe EJC-independent stimulation of degradation by tethered UPF3b. Furthermore, deletion or point mutants of UPF3b, which are unable to interact with UPF2, still induce mRNA degradation upon tethering to the transcript (Boehm et al., 2014; Gehring et al., 2003). Providing evidence from another angle, tethering of a MAGOH mutant which assembles into the EJC, but fails to interact with UPF3b because of disruption of the EBM binding site, still induces mRNA degradation (Gehring et al., 2005; Gehring et al., 2009a). As a side note, this MAGOH mutant is in general unable to interact with EBM-containing proteins, indicating that SMG6 with its two EBMs is either not recruited to the mRNA by tethered EJC components or that SMG6 is dispensable for EJC induced NMD (discussed below). In addition, degradation of EJC-triggered reporter mRNA was not inhibited by siRNA-mediated knockdown of UPF2 or UPF3b, which is consistent with the results obtained by tethering of individual EJC proteins in UPF2 depleted cells (Boehm et al., 2014; Gehring et al., 2005). Collectively, these observations challenge the current EJC-NMD model and suggest a UPF2- and UPF3b-independent manner of EJC-mediated NMD activation (Figure 9, steps 5 and 6). As currently no plausible alternative model for the mode of EJC activation is available, it is a matter of future research to elucidate the degradation inducing mechanism. One attractive possibility is that the EJC directly or indirectly interrupts the complicated interaction network that is required for proper translation

termination (Figure 9, step 5), although no experimental evidence for this hypothesis is available so far.

3.3 Comparison of EJC- and long 3′ UTR-induced NMD

Measured by the relative amount of endocleavage products, a long 3′ UTR is only a weak NMD stimulating element. In contrast, EJCs induce a very potent degradation of the reporter mRNA (Boehm et al., 2014). This difference in degradation potency could be connected to the likelihood to which these elements activate NMD at each upstream translation termination event. Although exact quantitative measurements of endocleavage events per translation cycle are not available and technically demanding, they could provide insights into the efficiency of different NMD-activating signals. Judged by qualitative analysis of the 3′ fragment to reporter mRNA ratio, long 3′ UTRs are unlikely to induce degradation at each translation termination event, whereas this could be more prevalent in the case of EJC-NMD. Recent temporal and spatial analysis of EJC-induced NMD showed that degradation via this pathway occurs most likely directly after export from the nucleus (Trcek et al., 2013). This could also explain the often-described dependence of NMD on the first round of translation, as it is conceivable that the first terminating ribosome upstream of an EJC is directly identified as premature. Therefore, EJCs seem to represent irrevocable signals for degradation, which not only ensure the rapid and efficient removal of erroneous transcripts, but also allow the regulated degradation of certain transcripts like SC35 for example.

In contrast, long 3′ UTRs are less defined signals and are probably regulated in many different ways due to their inhomogeneity. In general, they are weaker NMD inducers compared to EJCs, indicating that these elements are not specifically designed for the induction of degradation. The observation that long 3′ UTRs can induce NMD with varying efficiencies might be the result of an array of stochastic events occurring during translation termination, which are potentially determined by the UPF1 and PABPC1 occupancy closely around the stop codon.

Striking differences for the required components of the NMD machinery between both pathways were observed. EJC-NMD, as already discussed above, is not strictly dependent on UPF2 or UPF3b, whereas the degradation of long 3′ UTR NMD targets requires UPF2 (Boehm et al., 2014). This observation is further supported by complementation assays performed in UPF1 depleted cells. Whereas a UPF1 mutant unable to interact with UPF2 (VV204-205DI) can restore the degradation of EJC-induced NMD targets, the decay of TPI reporters with a long 3′ UTR is

not supported anymore (unpublished data). It is not clear yet, whether UPF3b is needed for this kind of NMD pathway, as the results obtained by complementation assays and tethering assays are not entirely congruent (Boehm et al., 2014). Interestingly, tethering of a UPF2 mutant which is unable to interact with UPF3b, results in over-stabilization of the reporter mRNA. This indicates that UPF3b is required to some extent for long 3′ UTR-activated NMD, which suggests that this NMD pathway is more similar to the one found in lower eukaryotes. Importantly, this implicates that EJCs and long 3′ UTRs do not only differ substantially from each other in the mode of NMD machinery activation, but also in the usage of NMD factors.

Dissecting the individual molecular reasons why NMD is either UPF2/UPF3b-independent or -dependent, is a challenging task for future research. Although we know multiple potential ways how UPF2 can influence UPF1, little is known about the role of UPF3b. So far, more in depth experiments with UPF3b were difficult, because UPF3a can substitute for UPF3b depletion. CRISPR-mediated knockout might help to solve this problem by the genomic inactivation of UPF3a, allowing for UPF3b complementation and functional assays. Concerning UPF2, it would be interesting to investigate which of its functions is required for long 3′ UTR-activated NMD. The first possibility is the modulation of the UPF1 helicase function, which could be mimicked by mutations on the interaction surface between the CH domain and the RecA2 domain of UPF1. Thereby, the CH domain cannot induce the conformational change required for clamping on the RNA and this simulates the UPF2-bound state to a certain extent (Chakrabarti et al., 2011). Secondly, UPF3b binding to UPF2 could somehow influence the activation of UPF1. If this effect is direct or involves the activation of, for example, SMG1 is unclear. Of interest, biochemical analysis obtained in this work, combined with the information from the UPF2-SMG1 cryo-EM structure, collectively shows that UPF2 interacts via the third MIF4G with the FRB domain of SMG1 (Clerici et al., 2014; Melero et al., 2014). Furthermore, the concurrent interaction of UPF2 with SMG1 and UPF3b was observed, indicating that both binding sites of the MIF4G-3 are not mutually exclusive (Clerici et al., 2014). Although we lack structural information about the UPF2-UPF3-SMG1 interaction on the atomic level, it is possible that UPF3b is bound to UPF2 in close proximity to SMG1 and, therefore, is in the spatial position that allows regulation of the SMG1 kinase activity. This is especially interesting, because as discussed above, mammalian long 3′ UTR-NMD exhibits characteristics similar to the evolutionary more ancient NMD pathway found for example also in *C. elegans*. For the NMD pathway in nematodes, UPF2 and UPF3 are required for UPF1 phosphorylation, suggesting that

these features are at least partially conserved in mammals (Page et al., 1999). To test this, specific point mutants disrupting the SMG1-UPF2 interaction need to be characterized in order to determine the impact of UPF2 on the UPF1 phosphorylation state.

After having obtained detailed structural information about the first two MIF4G domains of UPF2, the function of these domains was tested in complementation and tethering assays. Interestingly, both domains were required for normal NMD activity, while the UPF1/UPF3b binding and the cytoplasmic localization was unchanged by the deletions (Clerici et al., 2014). Combined with the data from the UPF-EJC cryo-EM structure, it was proposed that these domains likely serve a scaffolding role (Clerici et al., 2014; Melero et al., 2012). However, as conserved patches of the first MIF4G were found to be important for yeast NMD, it cannot be excluded that mammalian UPF2 MIF4G-1 and -2 are not also involved in critical, but so far unknown interactions (Fourati et al., 2014).

3.4 Degradation of the mRNA via endonucleolytic cleavage

Following the activation of NMD, the mRNA is ultimately degraded by potentially redundant pathways. These involve the accelerated deadenylation by SMG7 recruited CCR4-NOT deadenylase complex, the deadenylation-independent decapping via direct or PNRC2-mediated DCP1/2 activation, and the SMG6-catalyzed endonucleolytic cleavage (Figure 9, step 7). All these pathways have in common that the involved factors are supposedly directly recruited to the phosphorylated N- and C-terminus of UFP1. Consequently, phosphorylation of UPF1 represents the step in NMD upon which the degradation of the mRNA is ultimately initiated (Figure 9, steps 6 and 7). It is currently unclear, if these degradation pathways are entirely redundant, can co-exist with each other, or if specific pathways are used for certain mRNA substrates.

In Boehm et al. (2014), the requirements for SMG6-mediated endocleavage were intensively analyzed. Specific detection of SMG6 activity is possible due to the accumulation of 3′ endocleavage intermediates (3′ fragments) in XRN1 depleted cells, which arise only upon SMG6-mediated endonucleolytic cleavage. It was previously stated that SMG6 cleaves mRNA substrates around the stop codon (Eberle et al., 2009; Huntzinger et al., 2008). With nucleotide-resolution high-throughput sequencing we showed for several reporters and an endogenous NMD target that indeed most of SMG6 catalyzed endocleavage occurs at or downstream of the termination codon (Boehm et al., 2014). This implies that the ribosome has to be disassembled

or moved away from the termination site prior to SMG6-mediated endocleavage, as otherwise the region around the stop codon should be protected by the ribosome. Furthermore, this raises the question, whether only the UPF1 molecules which are located closest to the stalled ribosome are activated and phosphorylated, although UPF1 molecules are spread along the 3′ UTR. If UPF1 phosphorylation occurs position-independently, we would expect to observe endocleavage products at various downstream regions as well, which was not the case. In line with this hypothesis, RNA-immunoprecipitations showed that phosphorylated UPF1 preferentially associates with the 5′ end of the 3′ UTR (Kurosaki et al., 2014). Further transcriptome-wide nucleotide-resolution analysis of the phospho-UPF1 position on the mRNA could provide more information about this distribution and subsequently about the UPF1 activation mechanism.

In order to cleave the mRNA endonucleolytically, SMG6 first needs to be recruited to the mRNP. Although the interaction with the phosphorylated N-terminus of UPF1 via the 14-3-3-like domain of SMG6 is the currently accepted mechanism, two alternative recruitment pathways have been proposed. One involves the N-terminal EBMs of SMG6, which mediate the interaction with the EJC and were reported to be essential for NMD (Kashima et al., 2010). In contrast, SMG6 EBM point and deletion mutants, which were unable to interact with EJCs, were analyzed in complementation assays using varying reporter mRNA and it was determined that the EBMs are dispensable for the execution of endocleavage (Boehm et al., 2014). Therefore, it is currently unclear what the functions of the EBMs are, although it cannot be excluded that they contribute to certain NMD events, which require EJC-specific SMG6 loading.

The second alternative recruitment pathway was proposed based upon *in vitro* binding studies, showing that the unstructured region in the SMG6 N-terminus is sufficient to mediate the phosphorylation-independent binding to UPF1 (Chakrabarti et al., 2014). A second publication supported the phosphorylation-independent association of SMG6 with UPF1, suggesting that this mode of interaction indeed could have functional relevance (Nicholson et al., 2014). Since no stable interaction of the SMG6 14-3-3-like domain with hyperphosphorylated UPF1 was detected *in vitro*, it is currently unclear how SMG6 is recruited to UPF1. Even if SMG6 can directly interact with hypophosphorylated UPF1, this has to occur in a regulated manner in the cellular context, as otherwise SMG6 would constantly be recruited to all 3′ UTRs. The other

alternative is that SMG6 is recruited in an inactive state, which requires further progression in the NMD pathway to allow the activation of SMG6.

It is currently a matter of debate why NMD employs several seemingly redundant degradation pathways. The situation in mammalian cells might be the consequence of the gradual expansion of the number of NMD components during evolution, with the goal to gain more options for regulation. This is also reflected by the intensive usage of EJCs as NMD-inducing markers, since as a result the NMD pathway in mammals features a very potent way of targeting mRNAs for degradation. In contrast, long 3′ UTR-containing mRNAs are not supposed to be degraded as efficiently, suggesting that these mRNAs can be subjected to more intensive regulation. Of the potential degradation pathways, endocleavage around the stop codon represents arguably the most efficient mechanism, as the endonuclease SMG6 is in theory directly able to cleave the RNA once recruited to the target. The two generated RNA fragments are rapidly degraded due to the unprotected 5′ and 3′ ends and SMG6-mediated cleavage is therefore most likely an irreversible step of mRNA decay (Figure 9, step 8). Although the exosome was reported to be required for the 3′-5′ degradation of 5′ fragments in flies, these fragments could not be stabilized in mammalian cells depleted for core exosome components, suggesting that additional factors are involved in the degradation (Eberle et al., 2009; Huntzinger et al., 2008).

In contrast to SMG6, SMG5/7- and PNRC2-mediated induction of exonucleolysis requires the establishment of interaction cascades, which are not formed instantaneously. Furthermore, deadenylation-dependent decapping requires a series of events, starting with the shortening of the poly(A) tail, followed by the protection of the 3′ end of the mRNA by the Lsm1-7-Pat1 (Tharun, 2009; Totaro et al., 2011). This in turn activates the decapping complex, which recruits XRN1 that catalyzes the subsequent 5′-3′ degradation of the mRNA body (Braun et al., 2012). Of note, the activities of cytoplasmic polyadenylation and capping enzymes were described, suggesting that during the earlier steps of exonucleolytic decay the mRNA can potentially still be re-converted into a functional and translation-capable mRNP (Charlesworth et al., 2013; Kiss et al., 2015). In conclusion, endocleavage is a fast, efficient and likely irreversible way of eliminating transcripts, whereas exonucleolytic degradation from the 5′ and 3′ end of the mRNA requires multiple subsequent steps, is a more adjustable process and is potentially reversible.

Recent high-throughput sequencing experiments showed that SMG6-mediated endocleavage is the preferred NMD degradation pathway (Lykke-Andersen et al., 2014; Schmidt et al., 2014).

However, other studies showed that for strong NMD inhibition, the knockdown of either SMG6 or SMG7 alone is not sufficient, in consequence both need to be depleted for substantial reporter mRNA level increase (Jonas et al., 2013). It is still unknown, whether both pathways (initiated by SMG6 or SMG5/7) operate independently or are somehow connected and regulate each other. As the factors required for the initiation of degradation share the majority of their binding sites (N- and C-terminus of UPF1), it is conceivable that there is a cross-talk between SMG5-7, PNRC2 and/or DCP1/2. Interestingly, *in vitro* binding studies showed that due to the phosphorylation-independent interaction of SMG6 with UPF1, both SMG6 and SMG5/7 can in principle be accommodated simultaneously on phosphorylated UPF1 (Chakrabarti et al., 2014).

4. References

Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K., and Dreyfuss, G. (1986). mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Molecular and cellular biology *6*, 2932-2943.

Alexandrov, A., Colognori, D., Shu, M.D., and Steitz, J.A. (2012). Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. Proceedings of the National Academy of Sciences of the United States of America *109*, 21313-21318.

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature *432*, 112-118.

Amrani, N., Sachs, M.S., and Jacobson, A. (2006). Early nonsense: mRNA decay solves a translational problem. Nature reviews Molecular cell biology *7*, 415-425.

Anders, K.R., Grimson, A., and Anderson, P. (2003). SMG-5, required for C.elegans nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. The EMBO journal *22*, 641-650.

Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Seraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science *313*, 1968-1972.

Applequist, S.E., Selg, M., Raman, C., and Jack, H.M. (1997). Cloning and characterization of HUPF1, a human homolog of the Saccharomyces cerevisiae nonsense mRNA-reducing UPF1 protein. Nucleic acids research *25*, 814-821.

Aravind, L., and Koonin, E.V. (2000). Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. Genome research *10*, 1172-1184.

Arias-Palomo, E., Yamashita, A., Fernandez, I.S., Nunez-Ramirez, R., Bamba, Y., Izumi, N., Ohno, S., and Llorca, O. (2011). The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. Genes & development *25*, 153-164.

Aronoff, R., Baran, R., and Hodgkin, J. (2001). Molecular identification of smg-4, required for mRNA surveillance in C. elegans. Gene *268*, 153-164.

Azzalin, C.M., and Lingner, J. (2006). The human RNA surveillance factor UPF1 is required for S phase progression and genome stability. Current biology : CB *16*, 433-439.

Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B., and Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. Nature structural & molecular biology *12*, 861-869.

Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., and Le Hir, H. (2012). Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. Nature structural & molecular biology *19*, 983-990.

Baserga, S.J., and Benz, E.J., Jr. (1988). Nonsense mutations in the human beta-globin gene affect mRNA metabolism. Proceedings of the National Academy of Sciences of the United States of America *85*, 2056-2060.

Becker, T., Franckenberg, S., Wickles, S., Shoemaker, C.J., Anger, A.M., Armache, J.P., Sieber, H., Ungewickell, C., Berninghausen, O., Daberkow, I., *et al.* (2012). Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. Nature *482*, 501-506.

Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., and Izaurralde, E. (2007a). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. The EMBO journal *26*, 1591-1601.

Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Sauliere, J., Wittkopp, N., and Izaurralde, E. (2007b). mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS letters *581*, 2845-2853.

Belgrader, P., Cheng, J., and Maquat, L.E. (1993). Evidence to implicate translation by ribosomes in the mechanism by which nonsense codons reduce the nuclear level of human triosephosphate isomerase mRNA. Proceedings of the National Academy of Sciences of the United States of America *90*, 482-486.

Bensimon, A., Aebersold, R., and Shiloh, Y. (2011). Beyond ATM: the protein kinase landscape of the DNA damage response. FEBS letters *585*, 1625-1639.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nature reviews Genetics *15*, 163-175.

Beznoskova, P., Cuchalova, L., Wagner, S., Shoemaker, C.J., Gunisova, S., von der Haar, T., and Valasek, L.S. (2013). Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. PLoS genetics *9*, e1003962.

Bhattacharya, A., Czaplinski, K., Trifillis, P., He, F., Jacobson, A., and Peltz, S.W. (2000). Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. Rna *6*, 1226-1235.

Boehm, V., Haberman, N., Ottens, F., Ule, J., and Gehring, N.H. (2014). 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. Cell reports *9*, 555-568.

Bollig, F., Winzen, R., Gaestel, M., Kostka, S., Resch, K., and Holtmann, H. (2003). Affinity purification of ARE-binding proteins identifies polyA-binding protein 1 as a potential substrate in MK2-induced mRNA stabilization. Biochemical and biophysical research communications *301*, 665-670.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell *126*, 713-725.

Bono, F., and Gehring, N.H. (2011). Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA biology *8*, 24-29.

Braun, J.E., Truffault, V., Boland, A., Huntzinger, E., Chang, C.T., Haas, G., Weichenrieder, O., Coles, M., and Izaurralde, E. (2012). A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. Nature structural & molecular biology *19*, 1324-1331.

Brocke, K.S., Neu-Yilik, G., Gehring, N.H., Hentze, M.W., and Kulozik, A.E. (2002). The human intronless melanocortin 4-receptor gene is NMD insensitive. Human molecular genetics *11*, 331-335.

Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., Le Hir, H., and Conti, E. (2010). Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. Proceedings of the National Academy of Sciences of the United States of America *107*, 10050-10055.

Buchwald, G., Schussler, S., Basquin, C., Le Hir, H., and Conti, E. (2013). Crystal structure of the human eIF4AIII-CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. Proceedings of the National Academy of Sciences of the United States of America *110*, E4611-4618.

Buhler, M., Steiner, S., Mohn, F., Paillusson, A., and Muhlemann, O. (2006). EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. Nature structural & molecular biology *13*, 462-464.

Cali, B.M., Kuchma, S.L., Latham, J., and Anderson, P. (1999). smg-7 is required for mRNA surveillance in Caenorhabditis elegans. Genetics *151*, 605-616.

Carter, M.S., Doskow, J., Morris, P., Li, S., Nhim, R.P., Sandstedt, S., and Wilkinson, M.F. (1995). A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. The Journal of biological chemistry *270*, 28995-29003.

Casadio, A., Longman, D., Hug, N., Delavaine, L., Vallejos Baier, R., Alonso, C.R., and Caceres, J.F. (2015). Identification and characterization of novel factors that act in the nonsense-mediated mRNA decay pathway in nematodes, flies and mammals. EMBO reports *16*, 71-78.

Chakrabarti, S., Bonneau, F., Schussler, S., Eppinger, E., and Conti, E. (2014). Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. Nucleic acids research *42*, 9447-9460.

Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011). Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. Molecular cell *41*, 693-703.

Chamieh, H., Ballut, L., Bonneau, F., and Le Hir, H. (2008). NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nature structural & molecular biology *15*, 85-93.

Chan, W.K., Bhalla, A.D., Le Hir, H., Nguyen, L.S., Huang, L., Gecz, J., and Wilkinson, M.F. (2009). A UPF3-mediated regulatory switch that maintains RNA surveillance. Nature structural & molecular biology *16*, 747-753.

Chan, W.K., Huang, L., Gudikote, J.P., Chang, Y.F., Imam, J.S., MacLean, J.A., 2nd, and Wilkinson, M.F. (2007). An alternative branch of the nonsense-mediated decay pathway. The EMBO journal *26*, 1820-1830.

Chang, J.C., and Kan, Y.W. (1979). beta 0 thalassemia, a nonsense mutation in man. Proceedings of the National Academy of Sciences of the United States of America *76*, 2886-2889.

Charlesworth, A., Meijer, H.A., and de Moor, C.H. (2013). Specificity factors in cytoplasmic polyadenylation. Wiley interdisciplinary reviews RNA *4*, 437-461.

Chauvin, C., Salhi, S., Le Goff, C., Viranaicken, W., Diop, D., and Jean-Jean, O. (2005). Involvement of human release factors eRF3a and eRF3b in translation termination and regulation of the termination complex formation. Molecular and cellular biology *25*, 5801-5811.

Chazal, P.E., Daguenet, E., Wendling, C., Ulryck, N., Tomasetto, C., Sargueil, B., and Le Hir, H. (2013). EJC core component MLN51 interacts with eIF3 and activates translation. Proceedings of the National Academy of Sciences of the United States of America *110*, 5903-5908.

Chen, Y.H., Su, L.H., and Sun, C.H. (2008). Incomplete nonsense-mediated mRNA decay in Giardia lamblia. International journal for parasitology *38*, 1305-1317.

Cheng, Z., Muhlrad, D., Lim, M.K., Parker, R., and Song, H. (2007). Structural and functional insights into the human Upf1 helicase core. The EMBO journal *26*, 253-264.

Cheng, Z., Saito, K., Pisarev, A.V., Wada, M., Pisareva, V.P., Pestova, T.V., Gajda, M., Round, A., Kong, C., Lim, M*., et al.* (2009). Structural insights into eRF3 and stop codon recognition by eRF1. Genes & development *23*, 1106-1118.

Chiu, S.Y., Lejeune, F., Ranganathan, A.C., and Maquat, L.E. (2004). The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex. Genes & development *18*, 745-754.

Chiu, S.Y., Serin, G., Ohara, O., and Maquat, L.E. (2003). Characterization of human Smg5/7a: a protein with similarities to Caenorhabditis elegans SMG5 and SMG7 that functions in the dephosphorylation of Upf1. Rna *9*, 77-87.

Cho, H., Kim, K.M., and Kim, Y.K. (2009). Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and decapping complex. Molecular cell *33*, 75-86.

Clerici, M., Deniaud, A., Boehm, V., Gehring, N.H., Schaffitzel, C., and Cusack, S. (2014). Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2. Nucleic acids research *42*, 2673-2686.

Clerici, M., Mourao, A., Gutsche, I., Gehring, N.H., Hentze, M.W., Kulozik, A., Kadlec, J., Sattler, M., and Cusack, S. (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. The EMBO journal *28*, 2293-2306.

Clissold, P.M., and Ponting, C.P. (2000). PIN domains in nonsense-mediated mRNA decay and RNAi. Current biology : CB *10*, R888-890.

Cosson, B., Berkova, N., Couturier, A., Chabelskaya, S., Philippe, M., and Zhouravleva, G. (2002). Poly(A)-binding protein and eRF3 are associated in vivo in human and Xenopus cells. Biology of the cell / under the auspices of the European Cell Biology Organization *94*, 205-216.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Crick, F.H. (1958). On protein synthesis. Symposia of the Society for Experimental Biology *12*, 138-163.

Cui, Y., Hagan, K.W., Zhang, S., and Peltz, S.W. (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. Genes & development *9*, 423-436.

Culbertson, M.R., and Leeds, P.F. (2003). Looking at mRNA decay pathways through the window of molecular evolution. Current opinion in genetics & development *13*, 207-214.

Culbertson, M.R., Underbrink, K.M., and Fink, G.R. (1980). Frameshift suppression Saccharomyces cerevisiae. II. Genetic properties of group II suppressors. Genetics *95*, 833-853.

Czaplinski, K., Ruiz-Echevarria, M.J., Paushkin, S.V., Han, X., Weng, Y., Perlick, H.A., Dietz, H.C., Ter-Avanesyan, M.D., and Peltz, S.W. (1998). The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. Genes & development *12*, 1665-1677.

Denning, G., Jamieson, L., Maquat, L.E., Thompson, E.A., and Fields, A.P. (2001). Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein. The Journal of biological chemistry *276*, 22709-22714.

Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835-845.

Doma, M.K., and Parker, R. (2007). RNA quality control in eukaryotes. Cell *131*, 660-668.

Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. Current biology : CB *12*, 1060-1067.

Drummond, D.A., and Wilke, C.O. (2009). The evolutionary consequences of erroneous protein synthesis. Nature reviews Genetics *10*, 715-724.

Durand, S., and Lykke-Andersen, J. (2013). Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. Nature structural & molecular biology *20*, 702-709.

Eberle, A.B., Lykke-Andersen, S., Muhlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nature structural & molecular biology *16*, 49-55.

Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z., and Muhlemann, O. (2008). Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. PLoS biology *6*, e92.

Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Current opinion in structural biology *20*, 313-324.

Fatscher, T., Boehm, V., Weiche, B., and Gehring, N.H. (2014). The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay. Rna *20*, 1579-1592.

Fernandez, I.S., Yamashita, A., Arias-Palomo, E., Bamba, Y., Bartolome, R.A., Canales, M.A., Teixido, J., Ohno, S., and Llorca, O. (2011). Characterization of SMG-9, an essential component of the nonsense-mediated mRNA decay SMG1C complex. Nucleic acids research *39*, 347-358.

Fiorini, F., Boudvillain, M., and Le Hir, H. (2013). Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. Nucleic acids research *41*, 2404-2415.

Fourati, Z., Roy, B., Millan, C., Coureux, P.D., Kervestin, S., van Tilbeurgh, H., He, F., Uson, I., Jacobson, A., and Graille, M. (2014). A Highly Conserved Region Essential for NMD in the Upf2 N-Terminal Domain. Journal of molecular biology *426*, 3689-3702.

Franckenberg, S., Becker, T., and Beckmann, R. (2012). Structural view on recycling of archaeal and eukaryotic ribosomes after canonical termination and ribosome rescue. Current opinion in structural biology *22*, 786-796.

Franks, T.M., Singh, G., and Lykke-Andersen, J. (2010). Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay. Cell *143*, 938-950.

Frischmeyer, P.A., and Dietz, H.C. (1999). Nonsense-mediated mRNA decay in health and disease. Human molecular genetics *8*, 1893-1900.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Molecular cell *17*, 537-547.

Gardino, A.K., Smerdon, S.J., and Yaffe, M.B. (2006). Structural determinants of 14-3-3 binding specificities and regulation of subcellular localization of 14-3-3-ligand complexes: a comparison of the X-ray crystal structures of all human 14-3-3 isoforms. Seminars in cancer biology *16*, 173-182.

Gatfield, D., and Izaurralde, E. (2004). Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. Nature *429*, 575-578.

Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. The EMBO journal *22*, 3960-3970.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Molecular cell *20*, 65-75.

Gehring, N.H., Lamprinaki, S., Hentze, M.W., and Kulozik, A.E. (2009a). The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS biology *7*, e1000120.

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009b). Disassembly of exon junction complexes by PYM. Cell *137*, 536-548.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., and Kulozik, A.E. (2003). Y14 and hUpf3b form an NMD-activating complex. Molecular cell *11*, 939-949.

Ghosh, S., and Jacobson, A. (2010). RNA decay modulates gene expression and controls its fidelity. Wiley interdisciplinary reviews RNA *1*, 351-361.

Glavan, F., Behm-Ansmant, I., Izaurralde, E., and Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. The EMBO journal *25*, 5117-5125.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014). MOV10 Is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs. Molecular cell *54*, 573-585.

Grimson, A., O'Connor, S., Newman, C.L., and Anderson, P. (2004). SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. Molecular and cellular biology *24*, 7483-7490.

Gross, J.D., Moerke, N.J., von der Haar, T., Lugovskoy, A.A., Sachs, A.B., McCarthy, J.E., and Wagner, G. (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell *115*, 739-750.

Guan, Q., Zheng, W., Tang, S., Liu, X., Zinkel, R.A., Tsui, K.W., Yandell, B.S., and Culbertson, M.R. (2006). Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. PLoS genetics *2*, e203.

Gudikote, J.P., Imam, J.S., Garcia, R.F., and Wilkinson, M.F. (2005). RNA splicing promotes translation and RNA surveillance. Nature structural & molecular biology *12*, 801-809.

Hagiwara, M., and Nojima, T. (2007). Cross-talks between transcription and post-transcriptional events within a 'mRNA factory'. Journal of biochemistry *142*, 11-15.

Hall, G.W., and Thein, S. (1994). Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia. Blood *83*, 2031-2037.

He, F., and Jacobson, A. (1995). Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. Genes & development *9*, 437-454.

He, F., and Jacobson, A. (2001). Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. Molecular and cellular biology *21*, 1515-1530.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. Molecular cell *12*, 1439-1452.

Hinnebusch, A.G. (2014). The scanning mechanism of eukaryotic translation initiation. Annual review of biochemistry *83*, 779-812.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics *123*, 301-313.

Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell *143*, 379-389.

Holbrook, J.A., Neu-Yilik, G., Hentze, M.W., and Kulozik, A.E. (2004). Nonsense-mediated decay approaches the clinic. Nature genetics *36*, 801-808.

Hoshino, S., Imai, M., Kobayashi, T., Uchida, N., and Katada, T. (1999). The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. The Journal of biological chemistry *274*, 16677-16680.

Huang, L., Lou, C.H., Chan, W., Shum, E.Y., Shao, A., Stone, E., Karam, R., Song, H.W., and Wilkinson, M.F. (2011). RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. Molecular cell *43*, 950-961.

Hug, N., and Caceres, J.F. (2014). The RNA helicase DHX34 activates NMD by promoting a transition from the surveillance to the decay-inducing complex. Cell reports *8*, 1845-1856.

Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. Rna *14*, 2609-2617.

Hurt, J.A., Robertson, A.D., and Burge, C.B. (2013). Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. Genome research *23*, 1636-1650.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L.E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell *106*, 607-617.

Isken, O., and Maquat, L.E. (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes & development *21*, 1833-1856.

Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W., and Kulozik, A.E. (2008). Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. The EMBO journal *27*, 736-747.

Izumi, N., Yamashita, A., and Ohno, S. (2012). Integrated regulation of PIKK-mediated stress responses by AAA+ proteins RUVBL1 and RUVBL2. Nucleus *3*, 29-43.

Jack, K., Bellodi, C., Landry, D.M., Niederer, R.O., Meskauskas, A., Musalgaonkar, S., Kopmar, N., Krasnykh, O., Dean, A.M., Thompson, S.R*., et al.* (2011). rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. Molecular cell *44*, 660-666.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2012). Termination and post-termination events in eukaryotic translation. Advances in protein chemistry and structural biology *86*, 45-93.

Jakobsen, C.G., Segaard, T.M., Jean-Jean, O., Frolova, L., and Justesen, J. (2001). [Identification of a novel termination release factor eRF3b expressing the eRF3 activity in vitro and in vivo]. Mol Biol (Mosk) *35*, 672-681.

Jeong, H.J., Kim, Y.J., Kim, S.H., Kim, Y.H., Lee, I.J., Kim, Y.K., and Shin, J.S. (2011). Nonsense-mediated mRNA decay factors, UPF1 and UPF3, contribute to plant defense. Plant & cell physiology *52*, 2147-2156.

Johansson, M.J., He, F., Spatrick, P., Li, C., and Jacobson, A. (2007). Association of yeast Upf1p with direct substrates of the NMD pathway. Proceedings of the National Academy of Sciences of the United States of America *104*, 20872-20877.

Jonas, S., Weichenrieder, O., and Izaurralde, E. (2013). An unusual arrangement of two 14-3-3-like domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. Genes & development *27*, 211-225.

Joncourt, R., Eberle, A.B., Rufener, S.C., and Muhlemann, O. (2014). Eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay by two genetically separable mechanisms. PloS one *9*, e104391.

Kadlec, J., Guilligay, D., Ravelli, R.B., and Cusack, S. (2006). Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. Rna *12*, 1817-1824.

Kadlec, J., Izaurralde, E., and Cusack, S. (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. Nature structural & molecular biology *11*, 330-337.

Kashima, I., Jonas, S., Jayachandran, U., Buchwald, G., Conti, E., Lupas, A.N., and Izaurralde, E. (2010). SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. Genes & development *24*, 2440-2450.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes & development *20*, 355-367.

Keeling, K.M., and Bedwell, D.M. (2011). Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases. Wiley interdisciplinary reviews RNA *2*, 837-852.

Keeling, K.M., Wang, D., Conard, S.E., and Bedwell, D.M. (2012). Suppression of premature termination codons as a therapeutic approach. Crit Rev Biochem Mol Biol *47*, 444-463.

Kerr, T.P., Sewry, C.A., Robb, S.A., and Roberts, R.G. (2001). Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay? Hum Genet *109*, 402-407.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nature reviews Molecular cell biology *13*, 700-712.

Kervestin, S., Li, C., Buckingham, R., and Jacobson, A. (2012). Testing the faux-UTR model for NMD: analysis of Upf1p and Pab1p competition for binding to eRF3/Sup35p. Biochimie *94*, 1560-1571.

Khajavi, M., Inoue, K., and Lupski, J.R. (2006). Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. Eur J Hum Genet *14*, 1074-1081.

Kiss, D.L., Oman, K., Bundschuh, R., and Schoenberg, D.R. (2015). Uncapped 5' ends of mRNAs targeted by cytoplasmic capping map to the vicinity of downstream CAGE tags. FEBS letters *589*, 279-284.

Klaholz, B.P. (2011). Molecular recognition and catalysis in translation termination complexes. Trends in biochemical sciences *36*, 282-292.

Kong, C., Ito, K., Walsh, M.A., Wada, M., Liu, Y., Kumar, S., Barford, D., Nakamura, Y., and Song, H. (2004). Crystal structure and functional analysis of the eukaryotic class II release factor eRF3 from S. pombe. Molecular cell *14*, 233-245.

Kononenko, A.V., Mitkevich, V.A., Atkinson, G.C., Tenson, T., Dubovaya, V.I., Frolova, L.Y., Makarov, A.A., and Hauryliuk, V. (2010). GTP-dependent structural rearrangement of the eRF1:eRF3 complex and eRF3 sequence motifs essential for PABP binding. Nucleic acids research *38*, 548-558.

Kozlov, G., and Gehring, K. (2010). Molecular basis of eRF3 recognition by the MLLE domain of poly(A)-binding protein. PloS one *5*, e10169.

Kozlov, G., Trempe, J.F., Khaleghpour, K., Kahvejian, A., Ekiel, I., and Gehring, K. (2001). Structure and function of the C-terminal PABC domain of human poly(A)-binding protein. Proceedings of the National Academy of Sciences of the United States of America *98*, 4409-4413.

Kunz, J.B., Neu-Yilik, G., Hentze, M.W., Kulozik, A.E., and Gehring, N.H. (2006). Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. Rna *12*, 1015-1022.

Kurosaki, T., Li, W., Hoque, M., Popp, M.W., Ermolenko, D.N., Tian, B., and Maquat, L.E. (2014). A post-translational regulatory switch on UPF1 controls targeted mRNA degradation. Genes & development *28*, 1900-1916.

Kurosaki, T., and Maquat, L.E. (2013). Rules that govern UPF1 binding to mRNA 3' UTRs. Proceedings of the National Academy of Sciences of the United States of America *110*, 3357-3362.

Lai, T., Cho, H., Liu, Z., Bowler, M.W., Piao, S., Parker, R., Kim, Y.K., and Song, H. (2012). Structural basis of the PNRC2-mediated link between mrna surveillance and decapping. Structure *20*, 2025-2037.

Lasalde, C., Rivera, A.V., Leon, A.J., Gonzalez-Feliciano, J.A., Estrella, L.A., Rodriguez-Cruz, E.N., Correa, M.E., Cajigas, I.J., Bracho, D.P., Vega, I.E*., et al.* (2014). Identification and functional analysis of novel phosphorylation sites in the RNA surveillance protein Upf1. Nucleic acids research *42*, 1916-1929.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. The EMBO journal *20*, 4987-4997.

Le Hir, H., Izaurralde, E., Maquat, L., and Moore, M. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. The EMBO journal *19*, 6860-6869.

Lee, K.M., and Tarn, W.Y. (2013). Coupling pre-mRNA processing to transcription on the RNA factory assembly line. RNA biology *10*, 380-390.

Leeds, P., Peltz, S.W., Jacobson, A., and Culbertson, M.R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. Genes & development *5*, 2303-2314.

Leeds, P., Wood, J.M., Lee, B.S., and Culbertson, M.R. (1992). Gene products that promote mRNA turnover in Saccharomyces cerevisiae. Molecular and cellular biology *12*, 2165-2177.

Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. The EMBO journal *21*, 3536-3545.

Lejeune, F., Li, X., and Maquat, L.E. (2003). Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Molecular cell *12*, 675-687.

Lelivelt, M.J., and Culbertson, M.R. (1999). Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. Molecular and cellular biology *19*, 6710-6719.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proceedings of the National Academy of Sciences of the United States of America *100*, 189-192.

Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M., and Cheung, V.G. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. Science *333*, 53-58.

Li, T., Shi, Y., Wang, P., Guachalla, L.M., Sun, B., Joerss, T., Chen, Y.S., Groth, M., Krueger, A., Platzer, M*., et al.* (2015). Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. The EMBO journal.

Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. Genes & development *27*, 2125-2138.

Longman, D., Hug, N., Keith, M., Anastasaki, C., Patton, E.E., Grimes, G., and Caceres, J.F. (2013). DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and Caenorhabditis elegans. Nucleic acids research *41*, 8319-8331.

Longman, D., Plasterk, R.H., Johnstone, I.L., and Caceres, J.F. (2007). Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. Genes & development *21*, 1075-1085.

Lykke-Andersen, J. (2002). Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. Molecular and cellular biology *22*, 8114-8121.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2000). Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell *103*, 1121-1131.

Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., and Jensen, T.H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. Genes & development *28*, 2498-2517.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.

Maquat, L.E., and Li, X. (2001). Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. Rna *7*, 445-456.

Maquat, L.E., Tarn, W.Y., and Isken, O. (2010). The pioneer round of translation: features and functions. Cell *142*, 368-374.

Matsuda, D., Hosoda, N., Kim, Y.K., and Maquat, L.E. (2007). Failsafe nonsense-mediated mRNA decay does not detectably target eIF4E-bound mRNA. Nature structural & molecular biology *14*, 974-979.

Meaux, S., van Hoof, A., and Baker, K.E. (2008). Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail. Molecular cell *29*, 134-140.

Medghalchi, S.M., Frischmeyer, P.A., Mendell, J.T., Kelly, A.G., Lawler, A.M., and Dietz, H.C. (2001). Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. Human molecular genetics *10*, 99-105.

Melero, R., Buchwald, G., Castano, R., Raabe, M., Gil, D., Lazaro, M., Urlaub, H., Conti, E., and Llorca, O. (2012). The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3' end. Nature structural & molecular biology *19*, 498-505, S491-492.

Melero, R., Uchiyama, A., Castano, R., Kataoka, N., Kurosawa, H., Ohno, S., Yamashita, A., and Llorca, O. (2014). Structures of SMG1-UPFs complexes: SMG1 contributes to regulate UPF2-dependent activation of UPF1 in NMD. Structure *22*, 1105-1119.

Mendell, J.T., ap Rhys, C.M., and Dietz, H.C. (2002). Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. Science *298*, 419-422.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nature genetics *36*, 1073-1078.

Metze, S., Herzog, V.A., Ruepp, M.D., and Muhlemann, O. (2013). Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. Rna *19*, 1432-1448.

Metzstein, M.M., and Krasnow, M.A. (2006). Functions of the nonsense-mediated mRNA decay pathway in Drosophila development. PLoS genetics *2*, e180.

Miller, J.N., and Pearce, D.A. (2014). Nonsense-mediated decay in genetic disease: friend or foe? Mutat Res Rev Mutat Res *762*, 52-64.

Mitchell, S.F., and Parker, R. (2014). Principles and properties of eukaryotic mRNPs. Molecular cell *54*, 547-558.

Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. Science *309*, 1514-1518.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.

Moriarty, P.M., Reddy, C.C., and Maquat, L.E. (1998). Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA. Molecular and cellular biology *18*, 2932-2939.

Mort, M., Ivanov, D., Cooper, D.N., and Chuzhanova, N.A. (2008). A meta-analysis of nonsense mutations causing human genetic disease. Human mutation *29*, 1037-1047.

Muhlemann, O., and Jensen, T.H. (2012). mRNP quality control goes regulatory. Trends in genetics : TIG *28*, 70-77.

Muhlrad, D., and Parker, R. (1999). Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. Rna *5*, 1299-1307.

Muller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nature reviews Genetics *14*, 275-287.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends in biochemical sciences *23*, 198-199.

Neu-Yilik, G., Amthor, B., Gehring, N.H., Bahri, S., Paidassi, H., Hentze, M.W., and Kulozik, A.E. (2011). Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon. Rna *17*, 843-854.

Nicholson, P., Josi, C., Kurosawa, H., Yamashita, A., and Muhlemann, O. (2014). A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. Nucleic acids research *42*, 9217-9235.

Nicholson, P., Yepiskoposyan, H., Metze, S., Zamudio Orozco, R., Kleinschmidt, N., and Muhlemann, O. (2010). Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cellular and molecular life sciences : CMLS *67*, 677-700.

Nirenberg, M.W., and Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proceedings of the National Academy of Sciences of the United States of America *47*, 1588-1602.

Nott, A., Le Hir, H., and Moore, M.J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. Genes & development *18*, 210-222.

Obsil, T., and Obsilova, V. (2011). Structural basis of 14-3-3 protein functions. Seminars in cell & developmental biology *22*, 663-672.

Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P., and Ohno, S. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Molecular cell *12*, 1187-1200.

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., and Ohno, S. (2012). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. Nucleic acids research *40*, 1251-1266.

Page, M.F., Carr, B., Anders, K.R., Grimson, A., and Anderson, P. (1999). SMG-2 is a phosphorylated protein required for mRNA surveillance in Caenorhabditis elegans and related to Upf1p of yeast. Molecular and cellular biology *19*, 5943-5951.

Pillers, D.A., Fitzgerald, K.M., Duncan, N.M., Rash, S.M., White, R.A., Dwinnell, S.J., Powell, B.R., Schnur, R.E., Ray, P.N., Cibis, G.W*., et al.* (1999). Duchenne/Becker muscular dystrophy: correlation of phenotype by electroretinography with sites of dystrophin mutations. Hum Genet *105*, 2-9.

Pisarev, A.V., Hellen, C.U., and Pestova, T.V. (2007). Recycling of eukaryotic posttermination ribosomal complexes. Cell *131*, 286-299.

Pisarev, A.V., Skabkin, M.A., Pisareva, V.P., Skabkina, O.V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U., and Pestova, T.V. (2010). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. Molecular cell *37*, 196-210.

Ponting, C.P. (2000). Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. Trends in biochemical sciences *25*, 423-426.

Preis, A., Heuer, A., Barrio-Garcia, C., Hauser, A., Eyler, D.E., Berninghausen, O., Green, R., Becker, T., and Beckmann, R. (2014). Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. Cell reports *8*, 59-65.

Pulak, R., and Anderson, P. (1993). mRNA surveillance by the Caenorhabditis elegans smg genes. Genes & development *7*, 1885-1897.

Ramani, A.K., Nelson, A.C., Kapranov, P., Bell, I., Gingeras, T.R., and Fraser, A.G. (2009). High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective Caenorhabditis elegans. Genome biology *10*, R101.

Rebbapragada, I., and Lykke-Andersen, J. (2009). Execution of nonsense-mediated mRNA decay: what defines a substrate? Current opinion in cell biology *21*, 394-402.

Rehwinkel, J., Letunic, I., Raes, J., Bork, P., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. Rna *11*, 1530-1544.

Remenyi, A., Scholer, H.R., and Wilmanns, M. (2004). Combinatorial control of gene expression. Nature structural & molecular biology *11*, 812-815.

Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C., and Riha, K. (2012). Aberrant growth and lethality of Arabidopsis deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. Nucleic acids research *40*, 5615-5624.

Rodriguez-Navarro, S., and Hurt, E. (2011). Linking gene regulation to mRNA production and export. Current opinion in cell biology *23*, 302-309.

Roque, S., Cerciat, M., Gaugue, I., Mora, L., Floch, A.G., de Zamaroczy, M., Heurgue-Hamard, V., and Kervestin, S.

(2015). Interaction between the poly(A)-binding protein Pab1 and the eukaryotic release factor eRF3 regulates translation termination but not mRNA decay in Saccharomyces cerevisiae. Rna *21*, 124-134.

Rufener, S.C., and Muhlemann, O. (2013). eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. Nature structural & molecular biology *20*, 710-717.

Salas-Marco, J., and Bedwell, D.M. (2004). GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. Molecular and cellular biology *24*, 7769-7778.

Sasikumar, A.N., Perez, W.B., and Kinzy, T.G. (2012). The many roles of the eukaryotic elongation factor 1 complex. Wiley interdisciplinary reviews RNA *3*, 543-555.

Sato, H., Hosoda, N., and Maquat, L.E. (2008). Efficiency of the pioneer round of translation affects the cellular site of nonsense-mediated mRNA decay. Molecular cell *29*, 255-262.

Sauliere, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonqueze, O., Audic, Y., Paillard, L., Roest Crollius, H., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. Nature structural & molecular biology *19*, 1124-1131.

Savas, S., Tuzmen, S., and Ozcelik, H. (2006). Human SNPs resulting in premature stop codons and protein truncation. Hum Genomics *2*, 274-286.

Schmid, M., and Jensen, T.H. (2010). Nuclear quality control of RNA polymerase II transcripts. Wiley interdisciplinary reviews RNA *1*, 474-485.

Schmidt, S.A., Foley, P.L., Jeong, D.H., Rymarquis, L.A., Doyle, F., Tenenbaum, S.A., Belasco, J.G., and Green, P.J. (2014). Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. Nucleic acids research.

Schoenberg, D.R. (2011). Mechanisms of endonuclease-mediated mRNA decay. Wiley interdisciplinary reviews RNA *2*, 582-600.

Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

Schweingruber, C., Rufener, S.C., Zund, D., Yamashita, A., and Muhlemann, O. (2013). Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochimica et biophysica acta *1829*, 612-623.

Serin, G., Gersappe, A., Black, J.D., Aronoff, R., and Maquat, L.E. (2001). Identification and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein and Upf3 protein (Caenorhabditis elegans SMG-4). Molecular and cellular biology *21*, 209-223.

Shigeoka, T., Kato, S., Kawaichi, M., and Ishida, Y. (2012). Evidence that the Upf1-related molecular motor scans the 3'-UTR to ensure mRNA integrity. Nucleic acids research *40*, 6887-6897.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nature structural & molecular biology *19*, 594-601.

Silva, A.L., Ribeiro, P., Inacio, A., Liebhaber, S.A., and Romao, L. (2008). Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. Rna *14*, 563-576.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell *151*, 750-764.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annual review of biochemistry.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS biology *6*, e111.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annual review of biochemistry *76*, 23-50.

Skabkin, M.A., Skabkina, O.V., Hellen, C.U., and Pestova, T.V. (2013). Reinitiation and other unconventional posttermination events during eukaryotic translation. Molecular cell *51*, 249-264.

Sladic, R.T., Lagnado, C.A., Bagley, C.J., and Goodall, G.J. (2004). Human PABP binds AU-rich RNA via RNA-binding domains 3 and 4. Eur J Biochem *271*, 450-457.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.

Song, H., Mugnier, P., Das, A.K., Webb, H.M., Evans, D.R., Tuite, M.F., Hemmings, B.A., and Barford, D. (2000). The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. Cell *100*, 311-321.

Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. Rna *5*, 221-234.

Steckelberg, A.L., Boehm, V., Gromadzka, A.M., and Gehring, N.H. (2012). CWC22 connects pre-mRNA splicing and exon junction complex assembly. Cell reports *2*, 454-461.

Sun, X., Moriarty, P.M., and Maquat, L.E. (2000). Nonsense-mediated decay of glutathione peroxidase 1 mRNA in the cytoplasm depends on intron position. The EMBO journal *19*, 4734-4744.

Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J., and Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. The EMBO journal *20*, 1785-1796.

Tange, T., Nott, A., and Moore, M. (2004). The ever-increasing complexities of the exon junction complex. Current opinion in cell biology *16*, 279-284.

Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y., and Akimitsu, N. (2012). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA biology *9*, 1370-1379.

Taylor, D., Unbehaun, A., Li, W., Das, S., Lei, J., Liao, H.Y., Grassucci, R.A., Pestova, T.V., and Frank, J. (2012). Cryo-

EM structure of the mammalian eukaryotic release factor eRF1-eRF3-associated termination complex. Proceedings of the National Academy of Sciences of the United States of America *109*, 18413-18418.

Tharun, S. (2009). Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay? RNA biology *6*, 228-232.

Thein, S.L., Hesketh, C., Taylor, P., Temperley, I.J., Hutchinson, R.M., Old, J.M., Wood, W.G., Clegg, J.B., and Weatherall, D.J. (1990). Molecular basis for dominantly inherited inclusion body beta-thalassemia. Proceedings of the National Academy of Sciences of the United States of America *87*, 3924-3928.

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. The EMBO journal *17*, 3484-3494.

Toma, K.G., Rebbapragada, I., Durand, S., and Lykke-Andersen, J. (2015). Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. Rna.

Totaro, A., Renzi, F., La Fata, G., Mattioli, C., Raabe, M., Urlaub, H., and Achsel, T. (2011). The human Pat1b protein: a novel mRNA deadenylation factor identified by a new immunoprecipitation technique. Nucleic acids research *39*, 635-647.

Trcek, T., Sato, H., Singer, R.H., and Maquat, L.E. (2013). Temporal and spatial characterization of nonsense-mediated mRNA decay. Genes & development *27*, 541-551.

Uchida, N., Hoshino, S., Imataka, H., Sonenberg, N., and Katada, T. (2002). A novel role of the mammalian GSPT/eRF3 associating with poly(A)-binding protein in Cap/Poly(A)-dependent translation. The Journal of biological chemistry *277*, 50286-50292.

Unterholzner, L., and Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. Molecular cell *16*, 587-596.

Villa, N., Do, A., Hershey, J.W., and Fraser, C.S. (2013). Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. The Journal of biological chemistry *288*, 32932-32940.

Wang, W., Cajigas, I.J., Peltz, S.W., Wilkinson, M.F., and Gonzalez, C.I. (2006). Role for Upf2p phosphorylation in Saccharomyces cerevisiae nonsense-mediated mRNA decay. Molecular and cellular biology *26*, 3390-3400.

Wang, W., Czaplinski, K., Rao, Y., and Peltz, S.W. (2001). The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. The EMBO journal *20*, 880-890.

Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Monch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E., Nerlov, C., and Porse, B.T. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes & development *22*, 1381-1396.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Molecular cell *2*, 135-140.

Wen, J., and Brogna, S. (2010). Splicing-dependent NMD does not require the EJC in Schizosaccharomyces pombe. The EMBO journal *29*, 1537-1551.

Weng, Y., Czaplinski, K., and Peltz, S.W. (1996a). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. Molecular and cellular biology *16*, 5477-5490.

Weng, Y., Czaplinski, K., and Peltz, S.W. (1996b). Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. Molecular and cellular biology *16*, 5491-5506.

Wiegand, H.L., Lu, S., and Cullen, B.R. (2003). Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. Proceedings of the National Academy of Sciences of the United States of America *100*, 11327-11332.

Wilson, M.A., Meaux, S., and van Hoof, A. (2008). Diverse aberrancies target yeast mRNAs to cytoplasmic mRNA surveillance pathways. Biochimica et biophysica acta *1779*, 550-557.

Wittkopp, N., Huntzinger, E., Weiler, C., Sauliere, J., Schmidt, S., Sonawane, M., and Izaurralde, E. (2009). Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. Molecular and cellular biology *29*, 3517-3528.

Wittmann, J., Hol, E.M., and Jack, H.M. (2006). hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. Molecular and cellular biology *26*, 1272-1287.

Wurtmann, E.J., and Wolin, S.L. (2009). RNA under attack: cellular handling of RNA damage. Crit Rev Biochem Mol Biol *44*, 34-49.

Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu, R., Morita, T., Iwamatsu, A., Hachiya, T.*, et al.* (2009). SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. Genes & development *23*, 1091-1105.

Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. (2001). Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. Genes & development *15*, 2215-2228.

Yepiskoposyan, H., Aeschimann, F., Nilsson, D., Okoniewski, M., and Muhlemann, O. (2011). Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. Rna *17*, 2108-2118.

Yoine, M., Nishii, T., and Nakamura, K. (2006). Arabidopsis UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. Plant & cell physiology *47*, 572-580.

Zaher, H.S., and Green, R. (2009). Fidelity at the molecular level: lessons from protein synthesis. Cell *136*, 746-762.

Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E. (1998a). At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. Molecular and cellular biology *18*, 5272-5283.

Zhang, J., Sun, X., Qian, Y., and Maquat, L.E. (1998b). Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. Rna *4*, 801-815.

Zhao, W., Blagev, D., Pollack, J.L., and Erle, D.J. (2011). Toward a systematic understanding of mRNA 3' untranslated regions. Proc Am Thorac Soc *8*, 163-166.

Zund, D., Gruber, A.R., Zavolan, M., and Muhlemann, O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. Nature structural & molecular biology *20*, 936-943.

5. Summary

Eukaryotic gene expression consists of a series of events mediating the information flow from DNA via mRNA to protein. Cellular surveillance mechanisms exist to detect and eliminate erroneous mRNA in order to prevent the production of incorrect transcripts. Nonsense-mediated mRNA decay (NMD) targets mRNA for degradation, which terminate translation prematurely or incorrectly. Thereby, NMD prevents the synthesis of unfunctional or harmful peptides. Besides this quality control function, NMD also regulates the levels of many full-length protein encoding mRNA.

The messenger ribonucleoprotein (mRNP) architecture downstream of the stop codon is the main determinant for the initiation of the NMD pathway. Exon-junction complexes (EJCs) and long 3′ untranslated regions (UTRs) are known stimulators of NMD. EJCs are central components of the gene expression pathway and are deposited upon splicing on the mRNA. The exact mechanism how these mRNP elements induce NMD is unclear. Moreover, the series of molecular events ultimately leading to the degradation of the mRNA, as well as the precise interplay of NMD factors during this process, are not well defined.

In this cumulative work, several important steps in the NMD pathway were investigated. I could show that that for NMD suppression, an interaction cascade involving the eukaryotic release factor 3 (eRF3), the cytoplasmic poly(A) binding protein (PABPC1) and the cap-binding EIF4F complex component eIF4G is required. This suggests that efficient ribosome recycling is important for the normal termination of translation, which in turn prohibits the activation of NMD. To gain insight into the mode of EJC assembly during splicing, CWC22, an essential splicing component, was identified as the critical EJC loading factor. Remarkable differences were observed when comparing long 3′ UTRs and EJCs as NMD-inducing elements. These differences involved not only the efficiency of mRNA degradation, the mode of NMD activation, but also the requirements of NMD factors. These results indicate that EJCs are highly evolved mRNP markers, which utilize a specific mechanism to achieve efficient degradation of the target mRNA. In contrast, long 3′ UTRs influence the mRNP composition around the stop codon, thus impairing regular translation termination and leading to infrequent and less efficient mRNA degradation. In conclusion, this work illuminates multiple aspects of mammalian NMD and highlights the important missing pieces of information, which are to be uncovered by future research.

6. Zusammenfassung

Die eukaryotische Genexpression besteht aus einer Reihe von Abläufen, welche den Informationsfluss von der DNA über mRNA bis zum Protein vermitteln. Dabei existieren zelluläre Überwachungsmechanismen die fehlerhafte mRNA detektieren und eliminieren, um dadurch die Produktion von inkorrekten Transkripten zu verhindern. Nonsense-vermittelter mRNA Abbau (NMD) führt zu dem Abbau von mRNA welche die Translation frühzeitig oder inkorrekt terminieren. Dadurch unterbindet NMD die Synthese von funktionsunfähigen oder schädlichen Peptiden. Neben dieser Qualitätskontrollfunktion, reguliert NMD auch viele mRNA die für Volllänge-Proteine kodieren.

Die messenger-Ribonucleoprotein (mRNP) Architektur strangabwärts des Stopcodons ist ausschlaggebend für die Initiierung des NMD Abbauweges. Exon-junction Komplexe (EJC) und lange 3′ untranslatierte Bereiche (UTR) sind bekannte NMD Stimulatoren. EJCs sind zentrale Bestandteile der Genexpression und werden während des Spleißens auf der mRNA platziert. Der exakte Mechanismus wie diese mRNP Elemente NMD induzieren können ist unklar. Weiterhin sind der Ablauf der molekularen Ereignisse welche schlussendlich zum Abbau der mRNA führen, wie auch das präzise Zusammenspiel der NMD Faktoren während dieses Prozesses, nicht gut definiert.

In dieser kumulativen Arbeit wurden mehrere wichtige Schritte des NMD Abbauweges untersucht. Dabei konnte ich zeigen, dass eine Interaktionskaskade für die Unterdrückung von NMD notwendig ist, welche den eukaryotischen Terminationsfaktor 3 (eRF3), das zytoplasmatische poly(A) Bindeprotein (PABPC1) und eIF4G, ein Bestandteil des Kappe-bindenden eIF4F Komplexes, beinhaltet. Dies weist darauf hin, dass effizientes Recycling des Ribosoms wichtig ist für die normale Termination der Translation, welche wiederum NMD verhindert. Zudem wurde das essentielle Spleißprotein CWC22 als der entscheidende Faktor für die Assemblierung von EJCs während des Spleißens identifiziert. Bemerkenswerte Unterschiede wurden deutlich beim Vergleich von langen 3' UTR und EJCs als NMD induzierenden Elementen. Dies betrifft nicht nur die Effizienz des mRNA-Abbaus, die Art und Weise der NMD-Aktivierung, sondern auch die Anforderungen an NMD Faktoren. Diese Ergebnisse zeigen, dass EJCs hochentwickelte mRNP Marker sind, die einen spezifischen Mechanismus verwenden, um einen effizienten Abbau der Ziel-mRNA zu erreichen. Im Gegensatz dazu beeinflussen lange 3' UTRs die mRNP Zusammensetzung in der Nähe des

Stopcodons, wodurch die reguläre Translations-Termination beeinträchtigt wird und dadurch seltener und weniger effizienter mRNA-Abbau eingeleitet wird. Zusammenfassend beleuchtet diese Arbeit mehrere Aspekte des Säugetier NMDs und hebt wichtige Punkte heraus, welche durch zukünftige Forschung zu beantworten sind.

7. Author contribution

- 3′ UTR length and mRNP composition determine endocleavage efficiencies at termination codons

A detailed author contribution is given in the end of the publication. In brief, I designed the study with N.H.G., performed the experiments, was involved in the analysis and interpretation of all data and wrote the manuscript with J.U. and N.H.G.

- The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay

Together with T.F. and N.H.G., I designed the study. T.F. and I performed an equal share of experiments. I was involved in the analysis and interpretation of the data and wrote the manuscript with T.F. and N.H.G. B.W. carried out the *in vitro* interaction studies and was involved in the analysis and interpretation of the data.

- Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2

M.C. and S.C. designed, performed and analyzed the structural part of the publication. A.D. and C.S. designed, performed and analyzed the *in vitro* interaction studies. Together with N.H.G., I designed, performed and analyzed the *in vivo* studies.

- CWC22 Connects Pre-mRNA Splicing and Exon Junction Complex Assembly

Together with A.S. and N.H.G I designed the study concerning CWC22 as an essential splicing component. I performed these experiments and analyzed and interpreted the corresponding data together with the coauthors.

8. Acknowledgement

Bei meinem Doktorvater Herrn Dr. Niels Gehring möchte ich mich an erster Stelle ganz herzlich für die gemeinsame sehr produktive und lehrreiche Zeit bedanken. Nicht nur wissenschaftlich, sondern auch persönlich habe ich mich enorm durch unsere Zusammenarbeit weiterentwickelt. Dabei hat mich stets seine Professionalität, Mentalität und wissenschaftliche Hingabe inspiriert und dadurch begeistert, Wissenschaft mit Leidenschaft und Enthusiasmus zu betreiben.

Ich danke meiner Prüfungskommission für das Interesse an dieser Arbeit, Frau Prof. Dr. Ute Höcker für die Übernahme des Vorsitzes und Frau Prof. Dr. Karin Schnetz für die Übernahme der Zweitkorrektur. Bei Dr. Martina Rembold bedanke ich mich für den Beisitz bei der Disputation.

Für die finanzielle Unterstützung und weitere Hilfe nicht nur bei organisatorischen Anliegen bedanke ich mich bei der Graduate School for Biological Sciences und der International Graduate School in Development Health and Disease.

Bei den aktuellen und ehemaligen Kollegen von „Gehringia“ bedanke ich mich für die sehr entspannte und offene Arbeitsatmosphäre, sowie für alle Diskussionen, mittägliche Lektüren und außeruniversitären Aktionen. Ich habe viele von meinen Kollegen als gute Freunde gewonnen, viel von ihnen gelernt und bin froh die vergangen Jahre zusammen mit ihnen verbracht zu haben.

Weiterhin möchte ich meinen engen Freunden für die verbrachte Zeit außerhalb der Universität danken, welche mir eine solide Balance im Leben ermöglicht hat. Speziell möchte ich Sita Arjune dafür danken, dass Sie mich während des gesamten Studiums bis zur Promotion immer unterstützt und motiviert hat.

Meiner Familie, im speziellen meinen Großeltern, Eltern und Brüdern möchte ich einen besonders großen Dank aussprechen für die großartige Unterstützung an jeder Stelle meines universitären Werdegangs. Wohl wissend was ich ihr in all den Jahren zu verdanken habe, wäre diese Arbeit in dieser Form ohne meine Familie nie zustande gekommen.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit − einschließlich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie − abgesehen von oben angegebenen Teilpublikationen − noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von PD Dr. Niels H. Gehring betreut worden.

Köln, den 23. April 2015

Volker Böhm

Volker Böhm

Persönliche Angaben

Geburtsdaten	20.03.1987 in Bergisch Gladbach
Anschrift	Mauritiussteinweg 51 50676 Köln
Telefon	0163 6394241
E-Mail	boehmv@uni-koeln.de
Familienstand	ledig
Staatsangehörigkeit	deutsch

Ausbildung

WS 2009 – WS 2013	Biological Sciences Masterstudiengang an der Universität zu Köln Masterarbeit: Molecular analysis of SMG5 and SMG7 in nonsense-mediated mRNA decay Bei PD Dr. Niels H. Gehring, Institut für Genetik Abschluss: Master of Sciences (Note 1,0)
WS 2006 – SS 2009	Biologie Studium an der Universität zu Köln Schwerpunkte: Biochemie und Biotechnologie Bachelorarbeit: Functional characterization of the active site of Arabidopsis nitrate reductase Bei Prof. Dr. Günter Schwarz, Institut für Biochemie Abschluss: Bachelor of Science (Note 1,3)
1997 – 2006	Stadtgymnasium Köln-Porz Abschluss: Abitur (Note 1,5)

Auszeichnungen und Stipendien

2010 – 2013	IGSDHD PhD Fellowship
2009 – 2013	Fast Track Masters/Doctoral Programme
2009 – 2010	NRW Stipendium

Köln, den 23. April 2015